**The Dissertation Committee for Rea Martine Lardelli Certifies that this is the approved version of the following dissertation:**

# Spliceosome assembly and rearrangements: understanding how snRNPs are built and helicases function

**Committee:**

Scott Stevens, Supervisor

Dean Appling

Karen Browning

David Hoffman

Rick Russell

# Spliceosome assembly and rearrangements: understanding how snRNPs are built and helicases function

**by**

**Rea Martine Lardelli, B.S.**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**August 2010**

# Dedication

For my parents and Sebastian.

I am unspeakably thankful for my loving, encouraging parents, who have never stopped supporting me or believing in me even in the hardest of times. I am also indescribably thankful for Sebastian, who has instilled confidence in me, who makes me laugh, and who gives me strength. Together, my parents and Sebastian are at the heart of what inspires me, and this dissertation is in their honor.

## Acknowledgements

I would like to express great thanks to my graduate supervisor Scott Stevens, for the extensive support and guidance. I am very fortunate to have spent my graduate studies in Stevens Lab, where I have been continuously challenged to become a better scientist. I would also like to express sincere thanks to my Ph.D. committee, Dean Appling, Karen Browning, David Hoffman, and Rick Russell for the advice and ideas regarding both specific research projects and future career directions.

I am so thankful for all of the members of the Stevens laboratory: Jennifer Hennigan, Sujin Lee, Matthew Sorenson, John Oxford, Amir Alpert, Christine Cucinotta, Allyson Spence, and Al Makrell, for their friendship and scientific insight. Additionally, I would like extend sincere gratitude to the past members of the Stevens Lab, Grace Chen, Josh Combs, Adam Roth, Champ Gupton, Eric Montemayor, and Jason Williams, for their encouragement, friendship, and advice. Arlen Johnson and his lab at UT have been especially helpful over the years, and I am also very thankful for the advice from Whitney Yin. The proteomic analyses with Maria Person and Michelle Gadush in the Protein Core Facility at UT, and the collaboration with James Thompson and the Yates Lab at Scripps have all been crucial for much of my thesis work and for this, I am very thankful. Thank you also to Bill Cassidy for his encouragement over the years. I am also very thankful for the RNA Society, which has given me the opportunity to extend my leadership skills. Thank you also to Marianna Grenadier for help with thesis figures.

I would be lost without the friendship, laughter, and advice from Dawn Klein, and I cannot thank her enough for her optimism and support. I am also so thankful for my friendships with Samantha Croft and Angeline Lyon. I am so thankful for all of my friends and family for their love and support.

**Spliceosome assembly and rearrangements:**
**understanding how snRNPs are built and helicases function**

Publication No._____

Rea Martine Lardelli, Ph.D.
The University of Texas at Austin, 2010

Supervisor: Scott Stevens

Pre-mRNA splicing by the spliceosome requires the precise and regulated efforts of the five snRNAs (U1, U2, U4, U5, and U6) and numerous associated proteins. Following assembly and activation of the spliceosome, two consecutive reactions result in intron removal and exon ligation from pre-mRNA substrates. It has been established that several members of the DExH/D-box family of helicases act transiently on the spliceosome prior to the chemical steps to authorize the successive reactions by hydrolyzing ATP and consequently inducing structural rearrangements. While it has been suggested that these changes produced in the structure of the spliceosome result in optimal positioning of the reactive species, the mechanisms and products of these reorganizations remain uncharacterized.

The work presented here describes the genetic strategy for accumulating and purifying spliceosomes arrested *in vivo,* during the catalytic steps of the splicing cycle. Using these complexes, we have defined the components required to proceed through the first and second steps of splicing, in addition to the factors required for the release of the spliced message. Analysis of these functional, synchronized particles has also allowed us to define a function for Prp2p in initiating the first step of pre-mRNA splicing. Our data suggest that Prp2p may act in an ATP-independent manner to remodel the spliceosome prior to using its ATPase function to displace the SF3 complex. We propose that the SF3 complex, in addition to its role in identification of the branchpoint, also acts to sequester the reactive 2'OH of the branchpoint adenosine to prevent premature reactivity.

Following the two catalytic steps of the splicing cycle, the spliceosome must disassemble and recycle its snRNPs for further rounds of splicing. The essential U6 snRNP component Prp24p, mediates one of the early assembly events - the annealing between the U4 and U6 snRNAs. We have discovered that although Prp24p is essential for viability, its function(s) can be bypassed by overexpressing the U6 snRNA. Additionally, biochemical characterizations of various forms of the U4/U6 snRNP provide evidence that Prp24p must be released before other components of the U4/U6 snRNP are permitted to interact and facilitate tri-snRNP formation.

# Table of Contents

# List of Tables

# List of Figures

# List of Illustrations

# Chapter 1: Introduction

## 1.1 EUKARYOTIC GENE EXPRESSION

The discovery that eukaryotes carry split genes, in which the coding regions are not continuous but interrupted by stretches of non-coding sequences called introns, was one of the most important discoveries in molecular biology in the second half of the 20th century (Berget *et al*., 1977; Chow *et al*., 1977). For proper gene expression, these introns must be removed and the blocks of coding sequences, referred to as exons, are joined together in a process known as precursor-mRNA (pre-mRNA) splicing. The vast majority of human genes contain multiple introns, which are spliced both constitutively and alternatively, presenting a clear argument for necessity of regulated and precise processing of pre-mRNA.

From a metabolic standpoint, the purpose of introns is not self-evident, as splicing is an energy consuming process (Cooper *et al*., 2009). However, as the potential for splicing events increases, so does the complexity of an organism. For example, roughly 94% of human genes undergo alternative splicing, and many of these post-transcriptional processing events take place in a tissue-, signal-, or developmental stage- specific manner (Wang *et al*., 2008). Along with the conferred advantages of introns and alternative splicing comes the responsibility of regulating the machinery that recognizes and functions in the removal of introns. With major cellular efforts directed toward precise and controlled intron removal, cells become vulnerable to misregulation, which in higher eukaryotes often results in disease (Cooper *et al*., 2009). It has been estimated that ~50% of disease causing point mutations within the coding sequences affect splicing (Lopez-Bigas *et al*., 2005).

The classical representation of gene expression holds that DNA is transcribed into RNA, and RNA is translated into proteins. Though this is the foundation for production of functional molecules, cells have evolved numerous regulatory steps throughout gene expression, and many of these events not only take place at the RNA level, but also involve catalytic functions of several classes of RNA. The coding RNAs, known as mRNAs, serve as the templates for protein synthesis, while non-coding small nuclear RNAs (snRNAs) act in complex with protein to remove the introns from the pre-mRNA. snRNAs have also been proposed to play a catalytic role in the maturation of pre-mRNAs (Yean *et al*., 2000). Other non-coding RNAs including, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), and microRNAs (miRNAs), are also critical for proper gene expression.

Pre-mRNA splicing is dependent upon a ribonucleoprotein (RNP) complex known as the spliceosome. Within the spliceosome, the coordinated activities of five snRNAs and over one hundred proteins recognize splice sites with high fidelity and accurately remove introns. This dissertation will specifically focus on the biochemical interactions and conformational changes within the spliceosome during the splicing cycle. The following introduction will detail the individual components of the splicing machinery, their mechanism of action, and the role of splicing regulation in control of gene expression.

### 1.1.1 Splicing history

In 1977, Philip Sharp (Berget *et al*., 1977) and Richard Roberts (Chow *et al*., 1977) suggested a new mechanism for the formation of mRNA upon the observation of incomplete hydrogen bonding between an adenovirus RNA transcript and the gene from which it was transcribed. Their data led to the "split gene" model in which large intervening sequences are removed during the formation of a mature mRNA species. Later, Walter Gilbert termed the gene disruptions "introns" and the coding regions which ultimately ligate together "exons" (Gilbert, 1978). The discovery of intron removal from pre-mRNA begged the questions of how this process occurred and what, if any, cellular machinery was required.

#### *1.1.1.1 The discovery of snRNAs*

Even before the discovery of introns, a class of RNAs was identified first in metazoan cells, which became known for its unique base composition and termed U-RNAs (Muramatsu *et al*., 1966; Okamura and Busch, 1965). From total, purified RNA, the U-RNAs sedimented in the 4 – 8S region of a sucrose gradient and were significantly different from rRNA and tRNAs (Muramatsu *et al*., 1966). The first of the U-RNAs to be isolated and sequenced were U1, U2 and U3 (Hodnett and Busch, 1968). Characterizations of these showed that U1 and U2 were nucleoplasmic while the third, U3, was localized to the nucleolus. Eventually three other snRNAs were discovered, each of which showed presence within the nucleus. With five of six of the RNAs localizing to the nucleus, and each having a notably high uridylic acid content, these RNAs were termed the U snRNAs (Busch *et al*., 1982).

Studies directed toward determining the existence of snRNAs in *Saccharomyces cerevisiae* revealed a more perplexing situation, as it seemed that at least 24 different snRNAs existed in budding yeast (Riedel *et al*., 1986; Wise *et al*., 1983). Further experimentation showed that these yeast snRNAs were encoded by single genes and existed at ~1/1000 the abundance of their mammalian counterparts (Wise *et al*., 1983). Despite the seemingly large number of snRNAs found in yeast compared to metazoan cells, only six of the yeast snRNAs we found to be essential, and these were the RNAs orthologous to the metazoan U1 – U6 snRNAs.

#### *1.1.1.2 The discovery of small nuclear ribonucleoproteins (snRNPs)*

Closely trailing the discovery of small nuclear RNAs, came the finding that proteins associate specifically with each of these snRNAs. Enger and Walters recognized that upon protease treatment, the snRNAs migrated more slowly in glycerol gradients, indicating each may be complexed with a number of proteins (Enger and Walters, 1970). The first protein components assigned to U1-U5 snRNAs were the Sm core of proteins, which is a seven-membered ring that binds a specific sequence in those snRNAs (Lerner and Steitz, 1979). Later it was discovered that although U6 does not have a consensus Sm binding sequence, it is bound by a "Like Sm" (LSm) core of proteins (Achsel *et al*., 1999; Vidal *et al*., 1999). The Sm complex has been shown to be required, but not sufficient for snRNP biogenesis in the cytoplasm, and the presence of NLS-like motifs in several of the Sm proteins imply that the Sm proteins may function in nuclear localization (Bordonne, 2000).

#### *1.1.1.3 snRNP discovery in splicing*

In the laboratory of Joan Steitz, a discovery in 1980 suggested that snRNAs were involved in splicing (Lerner *et al*., 1980). This work showed complementarity between the 5' sequence of the U1 snRNA and the consensus sequence of the splice junction, leading to the proposal that snRNAs may form base paired interactions in order to facilitate splicing of intron-containing transcripts. Later it was discovered by the Guthrie laboratory that the U2 snRNP recognizes the branchpoint sequence by base pairing interaction (Parker *et al*., 1987).

#### *1.1.1.4 The splicing mechanism*

Following the discovery of introns, great efforts were dedicated to understanding the mechanism by which introns were removed from pre-mRNA substrates. Major breakthroughs from the laboratories of John Abelson (Domdey *et al*., 1984; Lin *et al*., 1985), Michael Green (Grabowski *et al*., 1984; Padgett *et al*., 1984; Ruskin *et al*., 1984), Philip Sharp (Grabowski *et al*., 1984; Padgett *et al*., 1984), and Michael Rosbash (Rodriguez *et al*., 1984) proposed that the splicing reaction proceeds by means of two sequential transesterification reactions producing two sets of products, one from each reaction (Illustration 1.1).

In the first reaction, the 2' hydroxyl from an adenosine residue, termed the branchpoint adenosine, found within the conserved intron branchpoint sequence, nucleophilically attacks the phosphate preceding the conserved intronic GU at the 5' splice site. This yields the products of the first transesterification: a freed 5' exon and a lariat intermediate. In the second transesterification, the phosphate following the conserved intronic AG at the 3' splice site is attacked by the 3'-OH of the 5' splice site to

ligate the two exons and release the intron in the form of a lariat. Although, intron removal from self-splicing group II introns and pre-mRNA substrates is mechanistically identical, pre-mRNA intron-containing transcripts lack the extensive secondary and tertiary structure that group II introns possess, which allows cleavage in *cis*. Instead, the splicing of pre-mRNA depends upon trans-acting machinery, known as the spliceosome (Brody and Abelson, 1985; Lin *et al*., 1987).

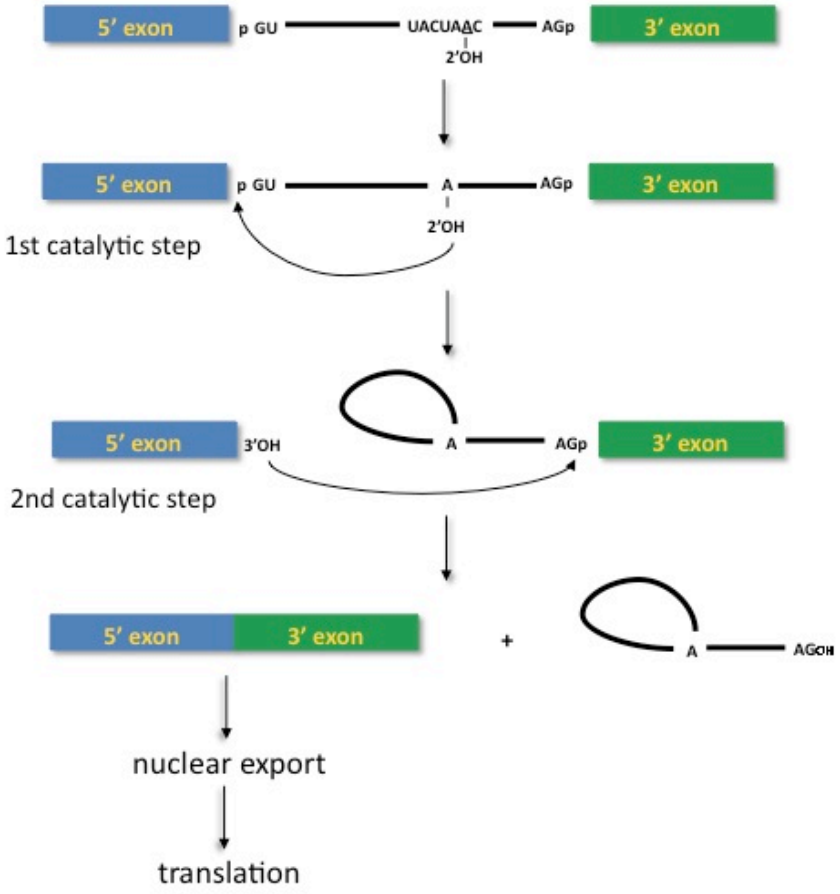


Illustration 1.1 Splicing proceeds through two stepwise transesterification reactions.

This illustration depicts the pre-mRNA, with the intron represented by the black line. The consensus 5' splice site, 3' splice site and branchpoint sequences are shown for *Saccharomyces cerevisiae*. In the first step of splicing the 2'OH from the branchpoint adenosine attacks the 5' SS to yield the lariat intermediate and the freed 5' exon. In the second catalytic step, the 5'SS attacks the 3' SS to form the spliced mRNA and the cleaved out lariat intron. Following splicing, the mRNA can be exported to the cytoplasm where it will be translated by the ribosome into protein.

#### *1.1.1.5 The spliceosome*

In humans, introns of up to several hundred kilobases can be processed and removed from pre-mRNA efficiently and accurately (Singh and Padgett, 2009). This suggests that eukaryotic cells have developed a mechanism for rapid and precise removal of introns of any size to ensure proper gene expression. The spliceosome positions the chemically reactive species in close enough proximity, overcoming the distances between the 5' splice site and the 3' splice site. This ribonucleoprotein complex collectively contains the five small nuclear RNAs and over one hundred stably-interacting and transiently-interacting proteins (Wahl *et al*., 2009). The yeast spliceosome was first characterized by the Abelson laboratory in 1985, when pre-mRNA and its splicing products were found to sediment at ~40S region upon glycerol gradient analysis (Brody and Abelson, 1985). This work employed yeast whole cell extract with radioactive *in vitro*-transcribed mutant pre-mRNA, and demonstrated splicing intermediate sedimentation corresponding to the intermediates produced from the two splicing reactions. Not long after the discovery of the spliceosome, it was found that numerous proteins, including snRNA-associated proteins, are involved in and required for commencement and completion of splicing. Because the spliceosome is a dynamic macromolecule, factors are being still identified and characterized to this day.

### 1.1.2 snRNP biogenesis

Because splicing is carried out in a cyclic pathway, and each snRNP enters into the splicing cycle in a unique manner, it is necessary to understand the origin of the snRNAs and their association with their repertoire of proteins.

### *1.1.2.1 U1, U2, U4 and U5 snRNP biogenesis*

Each of the snRNAs is transcribed by RNA polymerase II (RNA Pol II), except for U6 snRNA, which is transcribed by RNA Polymerase III (RNA Pol III). Possibly as a result of this difference, the biogenesis pathways for U1, U2, U4 and U5 snRNPs also differ significantly from that of the U6 snRNP (Kiss, 2004).

The snRNAs transcribed by RNA Pol II each undergo a number of processing events before they may participate in the splicing cycle. Following transcription of these snRNAs, the 5' end is modified with a 7-monomethyl guanosine ($m^7G$) cap. Like the other products of RNA Pol II transcription, the pre-snRNAs then interact with of the Cap Binding Complex (CBC) (Izaurralde *et al*., 1994), which acts as the adaptor for CRM1/RanGTP dependent nuclear export. An additional factor, Phosphorylated Adapter for RNA Export (PHAX), has been characterized as a specific requirement for snRNA export (Ohno *et al*., 2000). Following transport to the cytoplasm by means of the nuclear pore complex, the export machinery disassembles as triggered by the dephosphorylation of PHAX (Ohno *et al*., 2000). In higher eukaryotes, the survival of motor neurons (SMN) complex acts upon the snRNA by loading the heptameric core of Sm proteins onto the conserved binding site of the 3' end of each of the snRNAs (Massenet *et al*., 2002; Paushkin *et al*., 2002). The Sm complex, which is conserved from yeast through vertebrates, is composed of SmB1p, SmD1p, SmD2p, SmD3p, SmE1p, SmFp and SmGp, and is characteristic of each of the RNA Pol II transcribed snRNAs. This complex is essential for the subsequent steps of hypermethylation of the $^7$mG cap by Tgs1p methyltransferase and 3' end cleavage. Hypermethylation of the cap results in a 2,2,7-trimethylguanosine (TMG) cap, which along with the Sm complex, serves as the nuclear localization signal for snRNP reentry into the nucleus. A network of interactions between the cap, Snurportin-1 (SPN1), importin β, and a yet to be identified import factor act to

import the cytoplasmically processed snRNA (Palacios *et al*., 1997). Once back in the nucleus the snRNPs are further processed in sub-nuclear compartments, known as Cajal bodies in animals and plants. Here, the RNA component of the snRNP undergoes site-specific pseudouridylation and 2'-O-methylation (Sleeman and Lamond, 1999). Additionally, the Cajal bodies are also believed to be the sites of snRNP-specific protein association (Schaffert *et al*., 2004).

#### *1.1.2.2 U6 snRNP biogenesis*

The maturation pathway of the most highly conserved snRNA, U6 (Brow and Guthrie, 1988), differs significantly from the other snRNP processing events. To begin, the U6 snRNA receives a different 5' end modification with the addition of a γ-monomethyl cap (Shimba and Reddy, 1994), which probably contributes to U6 nuclear retention. As an RNA Pol III product, the U6 snRNA also has a 3'-oligouridine stretch, which is addressed by the La protein. The La protein is known to provide stability for the nascent transcript (Wolin and Cedervall, 2002). The 3' end of U6 eventually also possesses a 2', 3'-cyclic phosphate (Lund and Dahlberg, 1992), which is thought to trigger the release of the La protein and recruit the Like Sm (LSm) complex. The LSm ring is composed of seven proteins and binds to the 3' end of U6 as a heptameric ring, in a manner analogous to the Sm interaction with U1, U2, U4 and U5 (Achsel *et al*., 1999; Mayes *et al*., 1999; Seraphin, 1995; Vidal *et al*., 1999). Once the LSm complex associates, U6 snRNA is prevented from entering into the cytoplasm (Spiller *et al*., 2007). The LSm complex is also required to localize U6 snRNA to further processing within the nucleolus, where snoRNAs mediate the methylation and pseudouridylation of the snRNA (Ganot *et al*., 1999). Finally, and once again most likely directed by the LSm

complex, the U6 snRNA localizes to Cajal bodies (Gerbi and Lange, 2002). Here, U6 is thought to bind to its final snRNP protein, p110/SART3, (Stanek *et al*., 2003), which is orthologous to the yeast protein, Prp24p. Prp24p allows the U6 snRNP to form extensive base pairing interactions with the U4 snRNA (Shannon and Guthrie, 1991).

#### *1.1.2.3 Characteristic features of snRNP biogenesis in Saccharomyces cerevisiae*

Much less is known about snRNP biogenesis in *S. cerevisiae*. It remains unclear if snRNP maturation events in *S. cerevisiae* are restricted to the nucleus or, if U1, U2, U4, and U5, as in higher eukaryotes, go through a cytoplasmic phase. The Guthrie laboratory has identified a factor Brr1p, which shows association to each of the snRNPs and is distantly related to a factor contained in mammalian SMN complex (Noble and Guthrie, 1996). This relationship could imply a mechanism for yeast that is similar to Sm complex assembly in higher eukaryotes (Noble and Guthrie, 1996; Will and Luhrmann, 2001). Although no import factors analogous to Snurportin or importin β have been identified, evidence suggests that putative nuclear localization signals found within the carboxy-termini of SmB1p and SmD1p could function in recruiting the snRNPs back to the nucleus (Bordonne, 2000). Additional data from the Wolin laboratory suggests that the yeast homolog of the La protein, Lhp1p, may function not only in U6 stabilization and maturation, but also may protect the nascent U1, U2, U4, and U5 transcripts (Xue *et al*., 2000).

Each snRNA interacts with a number of snRNA-specific proteins in addition to the Sm complex or the LSm complex. Some of the snRNP-specific factors are found associated with their respective snRNAs and the spliceosome throughout the entire splicing cycle, while others interact in a more transient manner. In metazoan cells, these

factors are shown to load on to their respective snRNPs in Cajal bodies (Schaffert *et al*., 2004); however, the precise location of where these interactions initiate is not known in yeast and no compartment analogous to the Cajal body has been identified in *S. cerevisiae*.

### 1.1.3 The splicing cycle

Unlike most RNPs such as the ribosome, the spliceosome must assemble from its building blocks, the snRNPs, for each and every splicing event. As graphically depicted in Illustration 1.2, intron removal is executed only once the spliceosome is formed and remodeled by the addition and removal of proteins and formation and disruption of RNA-RNA interactions. Following the second catalytic step of splicing, the spliceosome must disassemble to recycle its snRNPs for further rounds of splicing. The dynamic and continuous cycle of the RNA-RNA and RNA-protein coordination within the spliceosome builds an RNP of unique and particular interest.

Classically, assembly of the spliceosome has been represented as the stepwise addition of snRNPs to the pre-mRNA, where the splice sites of the pre-mRNA recruit spliceosomal proteins and form base pairing interactions with snRNAs (Lin *et al*., 1985). This assembly pathway is represented in the assembly events shown in Illustration 1.2. Alternatively, the snRNPs have been shown to associate as a holoenzyme termed the penta-snRNP, which addresses the substrate only after all five snRNPs associate (Stevens *et al*., 2002). In either case, the snRNPs dynamically rearrange to ultimately interact with the substrate to form an active site where chemistry can take place. The beginning of this section will describe the traditional model for stepwise assembly of the spliceosome

starting with the recognition of the 5' splice site by the U1 snRNA. Following this review, the alternative penta-snRNP model will be presented.

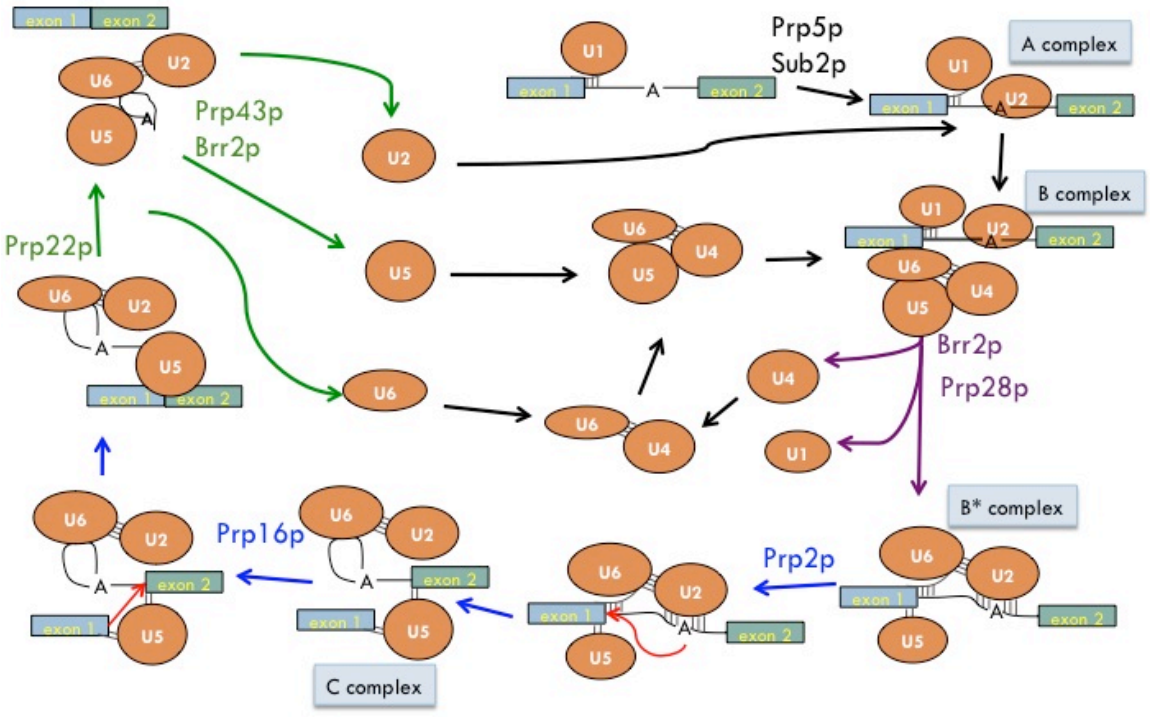


Illustration 1.2 Pre-mRNA splicing is a dynamic cycle, including assembly, activation, catalysis, disassembly, and recycling.

The pre-mRNA splicing cycle includes numerous dynamic rearrangements within the snRNP components, the pre-mRNA, and the assembled spliceosome. Many of the rearrangements are ATP-dependent. The yeast enzymes that carry out the ATP-dependent steps, termed RNA helicases, are highlighted throughout the cycle. The recycling and assembly steps of the splicing cycle are shown with black arrows and are dependent upon Prp5p and Sub2p. The activation steps are shown in purple, and are dependent on Brr2p and Prp28p. The two chemical steps are shown by blue arrows (flow of electrons in red), and are dependent upon Prp2p and Prp16p. The disassembly steps, including the message release, are shown with green arrows, and are dependent upon Prp22p, Prp43p and Brr2p. Adapted from Scott Stevens.

### *1.1.3.1 Commitment complex*

During transcription, the C-terminal domain (CTD) of RNA Pol II is responsible for recruiting the capping enzyme, which caps the 5' end of the nascent pre-mRNAs with a 7-methyl guanosine cap (Cho *et al*., 1997). Additionally, evidence suggests that the CTD is also responsible for co-transcriptional recruitment of the splicing machinery to the pre-mRNA (Fong and Bentley, 2001; Listerman *et al*., 2006). According to the stepwise model for spliceosome assembly, as the pre-mRNA is synthesized, the 5' end of the U1 snRNA and snRNP components recognize and interact with the 5' splice site in a ATP-independent manner. In yeast, the 5' splice site is the conserved dinucleotide, GU, at the 5' border of the intron (Reyes *et al*., 1996). This interaction forms what is known as the commitment complex (Ruby and Abelson, 1988; Seraphin and Rosbash, 1989). The RNA-RNA interaction is weak and believed to be aided by U1 snRNP components, likely including the following which have been shown to UV crosslink to the 5' splice site region: Snp1p, Yhc1p, Mud15p, Mud10p, SmB1p, SmD1p, and SmD3p (Puig *et al*., 1999; Zhang and Rosbash, 1999).

### *1.1.3.2 Pre-spliceosome formation*

Subsequent to the commitment complex formation, the Branchpoint Binding Protein, Bbp1p, recognizes the branchpoint (Berglund *et al*., 1998). In yeast the consensus branchpoint sequence is UACUAAC, where the underlined A is the branchpoint nucleotide. The 3' splice site consensus sequence is YAG, where Y is either pyrimidine. The 3' splice site is a variable distance downstream from the branchpoint sequence, but in yeast exists on average 39 nucleotides from of the branchpoint sequence (Spingola *et al*., 1999) and is separated by a polypyrimidine tract of 8-12 nucleotides.

This sequence, usually rich in uracil, is recognized by Mud2p (Zavanelli *et al*., 1994). Following the binding of Bbp1p to the branchpoint, the U2 snRNA recognizes and base pairs to the branchpoint sequence. It was later discovered that the branchpoint nucleotide is the only residue of the branchpoint sequence that does not pair with the U2 snRNA (Query *et al*., 1994), similar to the bulging branchpoint nucleotide of group II introns.

The ATP-dependent association of the U2 snRNP with the branchpoint requires the two DExH/D-box proteins, Sub2p and Prp5p (Fleckner *et al*., 1997; O'Day *et al*., 1996). The action of Sub2p may be required for the release of Bbp1p or Mud2p prior to U2 snRNP association with the pre-mRNA (Kistler and Guthrie, 2001), while Prp5p is likely involved in rearranging U2 snRNP prior to U2 interaction with the branchpoint (Abu Dayyeh *et al*., 2002). The interaction between the U2 snRNP and the branchpoint is dependent upon both the RNA-RNA interactions as well as RNA-protein interactions between the U2 snRNP components, SF3a and SF3b which bind upstream of the branchpoint and on the branchpoint, respectively (Gozani *et al*., 1996). The SF3a complex is composed of three factors in yeast, Prp9p, Prp11p and Prp21p (Ruby *et al*., 1993), while the SF3b complex is made up of six factors: Cus1p, Hsh49p, Hsh155p, Rse1p, Ysf3p, and Rds1p. Human pre-spliceosomes contain orthologs of the SF3 components, in addition to a factor, p14, which directly interacts with the branchpoint (Will *et al*., 2001), and to which there is no obvious yeast counterpart. This second recognition event for the branchpoint may be another checkpoint in the splicing cycle ensuring correct branchpoint choice. Once U1 and U2 are assembled upon the substrate, the spliceosome is termed the A complex.

#### *1.1.3.3 U4/U6•U5 formation*

According to the traditional stepwise model for assembly, a pathway parallel to the pre-spliceosome assembly exists to prepare U4/U6•U5, also referred to as tri-snRNP, for entry into the spliceosome. The final maturation of the mono-U6 snRNP involves the addition of the only mono-U6 snRNP specific protein, Prp24p (Shannon and Guthrie, 1991). As will be discussed later in detail, Prp24p has been shown to rearrange U6 snRNA such that U6 snRNA can anneal to U4 snRNA. Following its action, Prp24p dissociates, and U4 and U6 form an extensively base paired interaction. As shown in Illustration 1.3, this interaction extends the snRNAs into two intermolecular stems with one another (Brow and Guthrie, 1988). Additionally, other factors, Snu13p, Prp3p, Prp4p and Prp31p are known to bind to the U4/U6 snRNP, and a sequential model for their binding to di-snRNP has been suggested from human *in vitro* studies from the Lührmann laboratory (Nottrott *et al*., 2002). These factors are thought to aid in the stability of di-snRNP, as well as in the recruitment of mono-U5 to form tri-snRNP. As listed in Table 1.1, each of these factors is also found to be a component of tri-snRNP, aiding in the stability of this snRNP as well.

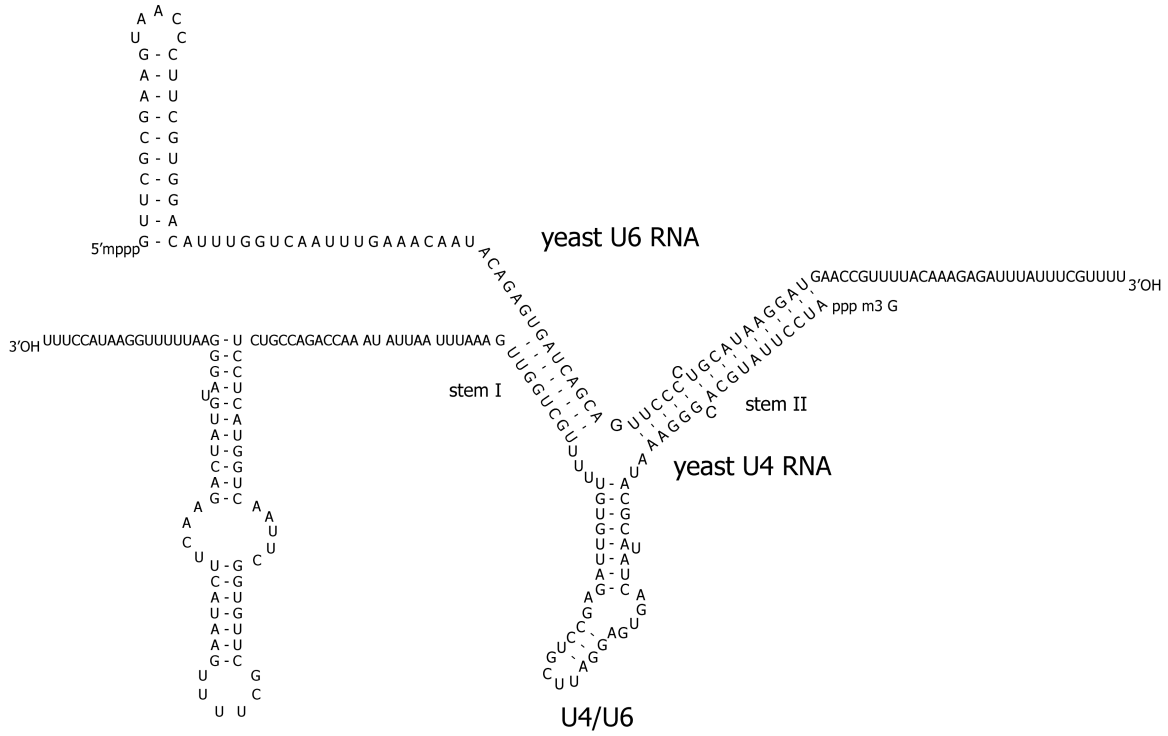


Illustration 1.3 U4 and U6 snRNAs form an extensive base pairing interaction upon di-snRNP formation.

The annealing between the two snRNAs is facilitated by the mono-U6 snRNP component, Prp24p. Prp24p likely facilitates this interaction by remodeling the structure of U6 snRNA into a conformation in which it can make the correct contacts for forming two intermolecular stems with U4 snRNA. U4 and U6 remain base paired throughout assembly, and upon activation, U4 snRNA is released from the spliceosome. Adapted from Jandrositz and Guthrie (1995) and Dunn and Rader (2010).

The core of the mono-U5 snRNP is composed of three large proteins, Brr2p, Prp8p and Snu114p, which interact with each other under high-salt conditions even in the absence of RNA (Achsel *et al*., 1998). This heterotrimeric complex likely functions in tri-snRNP formation, spliceosome formation, and spliceosome function, as it is present in the fully assembled spliceosome and subsists throughout the splicing cycle.

Human studies show how the interaction between U4/U6 and U5 snRNPs may be facilitated by interactions between the di-snRNP factors, Prp3p and Prp31p, and the tri-snRNP and mono-U5 snRNP factor, Prp6p (Liu *et al*., 2006; Makarova *et al*., 2002). Additionally, data support the tri-snRNP-specific component, Snu66p, may act to bridge di-snRNP and to mono-U5 snRNP through its protein-protein interactions with the U5 snRNP component Brr2p, and the di-snRNP component, Prp3p (Liu *et al*., 2006). The product of the extensive network of interactions between di-snRNP and mono-U5 is the formation of U4/U6•U5, or tri-snRNP. Tri-snRNP is composed of roughly 30 proteins, which are listed in Table 1.1.

| U1 snRNP | U2 snRNP | U5 snRNP | U6 snRNP | U4/U6 snRNP | U4/U6•U5 snRNP |
|---|---|---|---|---|---|
| Luc7p | Msl1p | Prp8p | Prp24p | LSm2p | Prp8p |
| Mud1p | Lea1p | Brr2p | LSm2p | LSm3p | Brr2p |
| Nam8p | Prp11p | Snu114p | LSm3p | LSm4p | Snu114p |
| Prp39p | Prp21p | Snu40p | LSm4p | LSm5p | Prp6p |
| Prp40p | Prp9p | Dib1p | LSm5p | LSm6p | Snu66p |
| SmB1p | Hsh49p | Prp6p | LSm6p | LSm7p | Prp3p |
| SmB2p | Hsh155p | SmB1p | LSm7p | LSm8p | Prp4p |
| SmD2p | Cus1p | SmB2p | LSm8p | SmB1p | Prp31p |
| SmD3p | Cus2p | SmD2p | | SmB2p | Snu13p |
| SmE1p | Ist3p | SmD3p | | SmD2p | Dib1p |
| SmFp | Res1p | SmE1p | | SmD3p | Prp38p |
| SmGp | Ysf3p | SmFp | | SmE1p | Spp381p |
| Snp1p | SmB1p | SmGp | | SmFp | Snu23p |
| Snu56p | SmB2p | Prp28p | | SmGp | SmB1p |
| Snu71p | SmD2p | | | Snu13p | SmB2p |
| Yhc1p | SmD3p | | | Prp3p | SmD2p |
| Prp42p | SmE1p | | | Prp4p | SmD3p |
| | SmFp | | | Prp31p | SmE1p |
| | SmGp | | | | SmFp |
| | Snu17p | | | | SmGp |
| | | | | | LSm2p |
| | | | | | LSm3p |
| | | | | | LSm4p |
| | | | | | LSm5p |
| | | | | | LSm6p |
| | | | | | LSm7p |
| | | | | | LSm8p |

Table 1.1 snRNP specific proteins

(Gottschalk *et al*., 2001a; Neubauer *et al*., 1997; Stevens and Abelson, 1999; Stevens *et al*., 2001)

### *1.1.3.4 B complex formation*

In the traditional model for assembly of the spliceosome, the B complex is defined by the presence of all five snRNAs and a pre-mRNA substrate. The composition of purified human B complex has been revealed by 2D gel analysis (Bennett *et al*., 1992) and by mass spectrometry (Makarov *et al*., 2002; Zhou *et al*., 2002). Within the B complex, RNA-RNA interactions are of considerable interest, as all five snRNAs are present. At the point in time when tri-snRNP joins the spliceosome, the 5' splice site is base paired with the U1 snRNA, while the branchpoint sequence is paired to the U2 snRNA. The U4 snRNA has not been shown known to interact with the pre-mRNA, while the U5 snRNP makes contacts with both the 5' splice site and the 3' splice site (Reyes *et al*., 1996; Umen and Guthrie, 1995a). The U6 snRNA undergoes extensive remodeling upon entry into the spliceosome; however, the significant rearrangements likely take place during activation (see 1.1.3.5 below).

Along with the addition of tri-snRNP to the spliceosome, a number of non-snRNP-associated proteins also join. Among these factors is the heteromeric Prp19p complex, known as the Nineteen Complex (NTC). This complex has been shown to be important for U5 snRNP and U6 snRNP stability during activation (Chen *et al*., 2002) and in U5 and U6 snRNA recognition of their respective binding sites in the events following activation (Chan and Cheng, 2005). Also known to bind at the time of B complex formation is the Retention and Splicing (RES) complex (Dziembowski *et al*., 2004). Although the B complex may contain all 5 snRNPs, it is not yet a catalytically active spliceosome.

### *1.1.3.5 Activation of the spliceosome: formation of B* complex*

Further conformational changes and compositional additions and removals are required to produce a splicing-competent spliceosome. It is thought that these rearrangements include, but are not limited to, the restructuring of the U6 snRNA. This is important to ensure the proper conformation for U6 snRNA and prevent premature cleavage of the substrate (Wahl *et al*., 2009). Activation of the B complex is traditionally characterized by the unwinding of U1 and U4 snRNAs from the spliceosome. The action on U4 is a result of the ATP-dependent unwinding by the DEID-box protein, Brr2p (Kim and Rossi, 1999; Lauber *et al*., 1996; Xu *et al*., 1996). Brr2p has multiple roles in splicing, and has been shown to be regulated by the GTPase protein Snu114p (Small *et al*., 2006). The ATP-dependent release of U1 is mediated by the DEAD-box factor, Prp28p (Staley and Guthrie, 1999). Upon activation, the U6 snRNA, which had previously been base paired to U4 snRNA, is now released from its interaction with U4 and exchanges with U1 snRNA interaction at the 5' splice site. Additionally, short duplexes between U2 snRNA and U6 snRNA are formed, and an intramolecular metal-binding stem-loop within the U6 snRNA forms (U6-ISL) (Yean *et al*., 2000). These conformational changes likely free residues within U6 that are important for catalysis. The remodeling events induced by Prp28p and Brr2p are depicted in Illustration 1.4, and specific RNA-RNA disruptions and interactions are highlighted. Prior to the start of this dissertation project, the activated "B*" complex had been the final spliceosome characterized prior to the first chemical step (Roth and Stevens, 2010, in preparation).

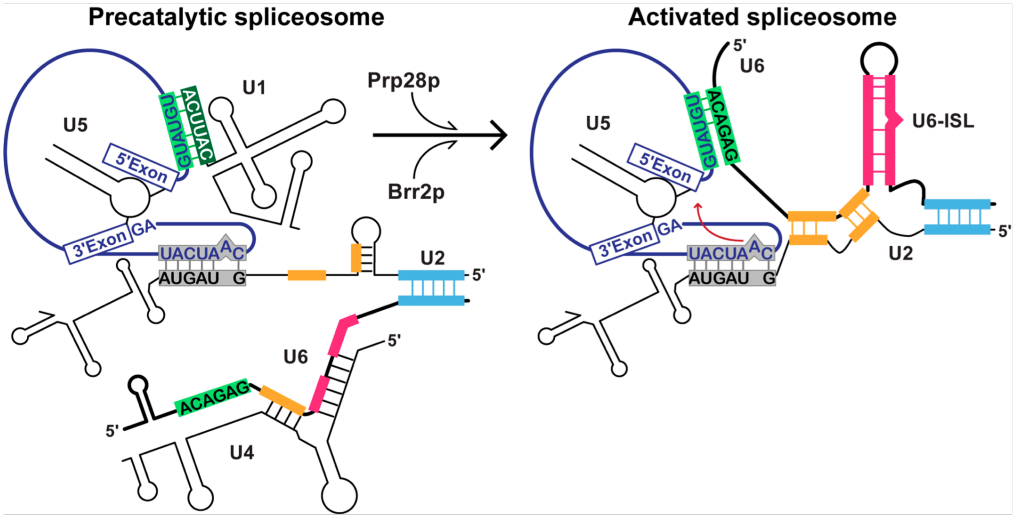


Illustration 1.4 Precatalytic spliceosomes (B complexes) dynamically rearrange to activate for catalysis.

Prp28p is known to unwind U1 snRNA from the 5' splice site (green), allowing for U6 snRNA to base pair to the 5' splice site (green). Brr2p releases U4 from the spliceosome, which results in the reformation of the U6-intra-stemloop (pink), in addition to short duplex formations between U2 and U6 (yellow). U2 snRNA remains base paired to the branchpoint sequence. Following the remodeling events, the spliceosome is activated; however, several other events, which are described in Chapter 3, are required before catalysis. Conserved yeast sequences of the pre-mRNA are depicted in dark blue. U1 and U6 snRNA/5' splice site-binding sequences are also shown. Adapted from Wahl *et al*. 2009.

#### *1.1.3.6 Transition from the B* complex to the C complex: the first chemical step*

The first step of splicing requires the DEAH-box protein, Prp2p (Chen and Lin, 1990; King and Beggs, 1990; Roy *et al*., 1995). Evidence supports a highly transient interaction of Prp2p with the spliceosome (King and Beggs, 1990), mediated by the G-patch protein, Spp2p (Roy *et al*., 1995). The hydrolysis of ATP by Prp2p is thought to cause a conformational rearrangement that results in the optimal positioning of the spliceosome around the pre-mRNA, creating an active site (Kim and Lin, 1996). However, the resulting structural composition and conformation of the rearrangement caused by Prp2p had only been speculated upon prior to this work. Recent work has suggested that another factor, Yju2p, acts in an ATP-independent manner after Prp2p hydrolysis, and prior to the first chemical step (Liu *et al*., 2007). Although prior to this work, activated B* complexes formed *in vivo* had not been compositionally defined, *in vitro* data exists to support major reorganizations within the spliceosome following the activation event (Makarov *et al*., 2002). Following these spliceosomal rearrangements, the 2' hydroxyl from the branchpoint adenosine attacks the 5' splice site to yield the lariat intermediate and the freed 5' exon.

#### *1.1.3.7 The C-complex*

Following the first step of splicing, the spliceosome is not competent for the next transesterification until it has undergone at least one more conformational change. To prepare for the second chemical step of splicing, the DEAH-box protein, Prp16p, hydrolyzes ATP, inducing a remodeling event which positions the reactive species, the 5' splice site and the 3' splice site, within close proximity (Schwer and Guthrie, 1992). Like Prp2p, Prp16p has only a transient interaction with the spliceosome, and following ATP

hydrolysis immediately dissociates (Schwer and Guthrie, 1991). Following this conformational change, several other factors have been characterized to play a major role in the preparation of the spliceosome for the second chemical reaction. These proteins include Prp18p and Slu7p, which have been shown to act after Prp16p but upstream of the chemical step (James *et al*., 2002). Additionally, Prp17p, which has been shown to join the spliceosome before the first step of chemistry, acts prior to the second step (Sapra *et al*., 2008). Finally the DEAH-box, Prp22p, has been shown to have an ATP-independent role in the second step of chemistry (Schwer and Gross, 1998). The actions of each of these proteins result in an active site where the 5' splice site can attack the 3' splice site, ligating the two exons and cleaving out the intron in the form of a lariat RNA.

#### *1.1.3.8 Post-catalysis*

Following the second chemical step of splicing, the spliceosome now contains the spliced mRNA and the lariat intron. Both of these RNAs are removed from the spliceosome by the actions of two DEAH-box factors. Prp22p, uses its ATP-dependent function to unwind the spliced message from the spliceosome (Schwer and Gross, 1998), while Prp43p is required to disassemble the post-splicing spliceosome, including the release of the lariat intron (Martin *et al*., 2002). The lariat is then debranched by Dbr1p (Chapman and Boeke, 1991) and is later degraded.

### *1.1.3.9 Preformed model for assembly*

Though strong evidence exists for the classical model for spliceosome assembly taking place in a stepwise manner, purifications of the snRNPs under physiological salt conditions suggest an alternative model for spliceosome formation.

In 1999, Stevens and Abelson characterized a 25S tri-snRNP particle under a salt concentration of 250 mM (Stevens and Abelson, 1999). Although tri-snRNP was isolated from cells and stable at 250 mM salt, the authors observed that when the salt concentration was lowered to 150 mM, U2 snRNP associated with the purified 30S particle, indicating the presence of a tetra-snRNP. The results were consistent with spliceosomes formed from human extracts conditions containing U2, U4, U5 and U6 (Konarska and Sharp, 1988). Konarska named the tetra-snRNP a pseudospliceosome and used the model to describe spliceosome formation as being a network of snRNP-snRNP interactions. Next, Stevens *et al*. lowered the salt to 50 mM, or conditions required for *in vitro* pre-mRNA splicing (Lin *et al*., 1985) and observed that all five snRNAs co-purified in a 45S particle in the absence of a pre-mRNA (Stevens *et al*., 2002). The presence of all five snRNPs and the absence of the pre-mRNA led to the alternative proposal that spliceosome assembly is carried out when the 45S "holo-enzyme" binds to a substrate, rather than assembled by discrete integrations of snRNPs upon a substrate. This model is outlined in Illustration 1.5.

The penta-snRNP model for assembly specifies that the U1 snRNA base pairs with the 5' splice site of the pre-mRNA and the branchpoint is base paired to the U2 snRNA all in the context of the other snRNPs. Furthermore, the data support that in some events, U1 snRNP commits to the pre-mRNA in the absence of the other snRNPs. As shown in Illustration 1.5, the tetra-snRNP, containing U2, U4, U5, and U6, is then

recruited. In this event, the U1 snRNP may be "marking" the 5' splice site to ensure proper processing of the substrate. In either case, the recruitment of a penta-snRNP to the substrate would be a kinetic advantage to gene expression. Finally, it is important to note the rearrangements required to produce a splicing competent spliceosome all take place in the context of the penta-snRNP, as all proteins previously described during spliceosome assembly were identified in the penta-snRNP, and *in vitro* splicing reactions confirmed the functionality of the penta-snRNP (Stevens *et al*., 2002).

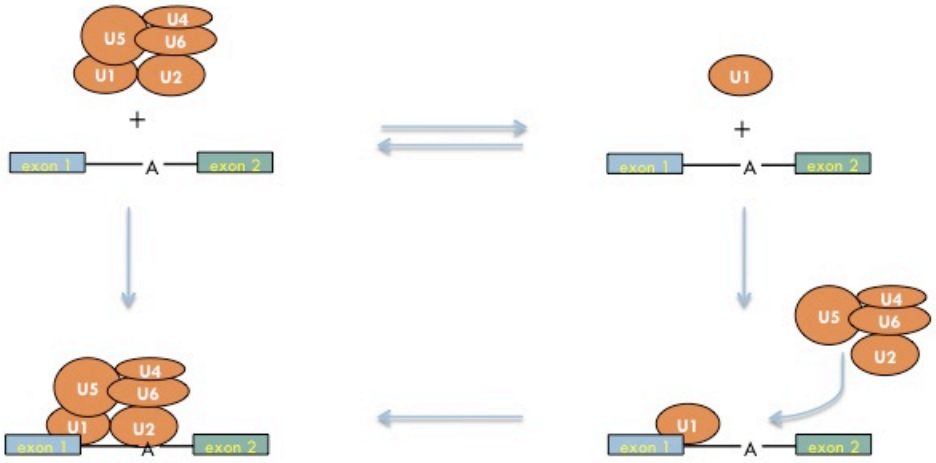


Illustration 1.5 The preformed model of spliceosome assembly states that a penta-snRNP contacts the pre-mRNA.

In the penta-snRNP model for assembly, all five snRNPs associate in the absence of pre-mRNA, as opposed to the stepwise addition of the snRNPs in assembly. In some cases U1 snRNP identifies the 5' splice site and then recruits the tetra-snRNP. In either case, the result is a B complex spliceosome, containing an unspliced pre-mRNA and all five snRNPs. In both models for assembly, the remodeling events are required to activate the spliceosome for catalysis. Adapted from Stevens *et al*. 2002.

Following the characterization of the yeast penta-snRNP, mammalian data was produced that supported the penta-snRNP assembly model (Azubel *et al*., 2004; Malca *et al*., 2003). Further work in human and chicken cells has described the existence of a supraspliceosome, where four penta-snRNPs associate upon a single substrate pre-mRNA (Azubel *et al*., 2006; Chen *et al*., 2007).

It is important to note that much of the data supporting stepwise assembly have been acquired from *in vitro* reconstitutions of the spliceosome or spliceosome assembly intermediates. In light of the salt-sensitivity of spliceosome precursors, one could argue that *in vitro*-reconstituted particles may not necessarily be authentic representations of the on-pathway intermediates of the splicing cycle. Additionally, *in vitro* assembly of the spliceosome produces ensemble populations, which are not well coordinated and which contain significant proportions of off-pathway, dead-end particles. Since the splicing cycle is just that, a cycle, purifications of particles arrested at specific steps *in vivo* may provide a more accurate description of the spliceosomal complexes that exist during the assembly, activation, catalysis, disassembly, and recycling.

## 1.2 DYNAMIC REARRANGEMENTS IN THE SPLICING CYCLE

Self-splicing group II introns depend on extensive secondary and tertiary structures within the intron to facilitate their intron removal (Boudvillain *et al*., 2000). Although pre-mRNA splicing uses the identical chemical mechanism for removal of introns, pre-mRNAs lack the intramolecular interactions and conserved sequences that group II introns employ to catalyze their intron removal. Instead pre-mRNAs depend on the actions of the snRNAs and proteins of the spliceosome to position the intron into a configuration where intron removal is achievable. This leads to the important question of how the dynamic changes within the spliceosome occur to promote spliceosome assembly, activation, catalysis and disassembly, all in a forward direction. The major contributors to this critical element of pre-mRNA processing are members of the DExH/D-box family of RNA helicases.

### 1.2.1 RNA helicases

RNA helicases are characterized by the presence of at least eleven identified conserved sequence motifs, which are involved in binding an NTP and coupling energy released by NTP hydrolysis to unwind double stranded RNA (de la Cruz *et al*., 1999). However, some RNA helicases have been implicated in other energy-dependent processes, which do not involve an unwinding event, but instead, an RNA-protein disruption or rearrangement (Jankowsky *et al*., 2001; Linder, 2006; Schwer, 2001). The family of RNA helicases that is responsible for most RNP modifications is known as the DExH/D-box helicases. The conserved sequences of DExH/D-box factors are depicted in Illustration 1.6. These factors have a unique sequence in motif II (DEAD or DEAH) of

the conserved helicase domain, for which they are named (Linder *et al*., 1989). DExH/D-box proteins have been implicated in almost all gene expression events, including transcription (Nakajima *et al*., 1997; Venema and Tollervey, 1999), ribosome biogenesis (Venema and Tollervey, 1999), pre-mRNA splicing (Staley and Guthrie, 1998), mRNA nuclear export (Schmitt *et al*., 1999), translation (Svitkin *et al*., 2001; Weng *et al*., 1996) and RNA degradation (Weng *et al*., 1996).

The differences between DEAD- and DEAH-box factors extend beyond their sequences in motif II. Although both are stimulated to hydrolyze an NTP by binding RNA, DEAD-box proteins use ATP exclusively, while DEAH-box factors may use any NTP (Tanaka and Schwer, 2005; Tanner *et al*., 2003). Additionally, DEAD-box factors contain two conserved sequences C-terminal to the DEAD-box domain (Cordin *et al*., 2004), while DEAH-box factors generally have longer conserved motifs within the DEAH domain.

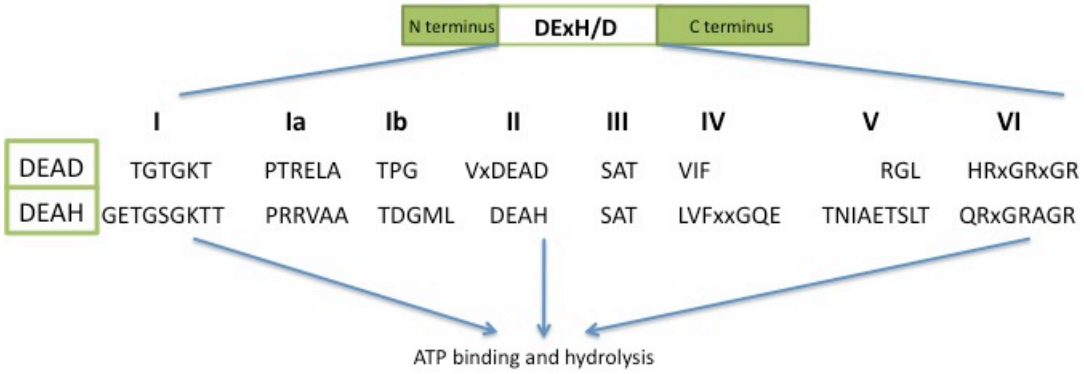


Illustration 1.6 Conserved motifs of DExH/D box proteins.

DExH/D-box proteins are defined by eight conserved motifs. Motifs I, II and VI are known to participate in ATP binding and hydrolysis, while motifs Ia, Ib, IV and V facilitate RNA binding. The DEAD-box motifs are typically shorter stretches of conserved sequences compared that of to DEAH-box proteins.

### 1.2.2 DExH/D-box helicases in splicing

The major conformational changes within the spliceosome are triggered by members of the DExH/D-box family. In the yeast splicing cycle, three DEAD-box factors are required in an ATP-dependent manner during spliceosome assembly: Sub2p (Kistler and Guthrie, 2001), Prp5p (Ruby *et al*., 1993), and Prp28p (Noble and Guthrie, 1996; Staley and Guthrie, 1999). In addition to these factors, there are four ATP-dependent steps during the catalytic and disassembly phases of the splicing cycle where DEAH-box proteins act sequentially to induce remodeling events. These factors include Prp2p, which is required before the first step of chemistry (King and Beggs, 1990; Schwer and Guthrie, 1991); Prp16p, which induces a rearrangement after the first step of chemistry but before the second step of chemistry (Schwer and Gross, 1998; Schwer and Guthrie, 1991); Prp22p, which uses its ATP-dependent function to release the spliced message from the spliceosome (Martin *et al*., 2002; Schwer and Gross, 1998); and Prp43p which disassembles the spliceosome after catalysis (Martin *et al*., 2002). Prp43p also functions in ribosome biogenesis (Combs *et al*., 2006; Lebaron *et al*., 2005; Leeds *et al*., 2006; Tanaka and Schwer, 2006). Additionally, Brr2p, which has two RNA helicase cores making it a hybrid between DEAD- and DExH-box families, has been shown to act in both activation (Raghunathan and Guthrie, 1998a) and disassembly of the spliceosome (Small *et al*., 2006). It is interesting to note that the helicases, which are required for catalysis during the splicing cycle are DEAH-box factors, while the assembly steps are promoted by DEAD-box proteins.

### 1.2.3 DEAH-box specificity

Although, most of the DExH/D-box factors have been shown to function in particular events, the specific targets of only three DExH/D-box factors have been determined (Guenther and Jankowsky, 2009). Among these, the specificity for only one splicing helicase, Prp22p, has been identified (Schwer, 2008). It is presumed that the reason for this paucity of information regarding helicase target sites is due to the innate dynamic characteristics of the helicases, creating an experimentally challenging problem. Many helicases, such as Prp2p and Prp16p, interact with their substrates in a transient manner probably to limit activity in time and space (Linder, 2006), thus making it difficult to determine interaction specificity. A recent biochemical strategy, Crosslinking analysis of cDNA (CRAC) (Granneman *et al*., 2009), which stems from Crosslinking Immunoprecipitation (CLIP) (Jensen and Darnell, 2008), has made it possible to identify the site on an RNA where an interacting protein binds. This strategy has already proven useful in identifying the interaction sites of RNA binding proteins (Bohnsack *et al*., 2009); however, it will be challenging for those helicases that interact only transiently in nature. Work in Chapter 2 and insights in Chapter 5 will elaborate upon strategies to identify binding sites for dynamically interacting RNA helicases.

### 1.2.4 DEAH-box-induced rearrangements

Each of the NTPase activities of DEAH-box factors has been shown to be stimulated by RNA binding. Recent work has revealed the first structure of the DEAH-box protein, Prp43p (He *et al*., 2010; Walbott *et al*., 2010). Structural work has revealed a β-hairpin between motifs V and VI, which the authors have proposed plays an important role in RNA unwinding or RNP remodeling. When bound by ADP, the β-hairpin of

Prp43p blocks access of an RNA substrate, indicating that hydrolysis of ATP or $P_i$ release triggers of remodeling by Prp43p (He *et al*., 2010). This finding may apply for the other DEAH-box factors as well, as many may process, translocate, or dissociate from their substrates concomitantly with ATP hydrolysis. The data also provide a consistent view of how structural rearrangements induced by the DEAH-box helicases are unidirectional.

Unfortunately, it is still unclear exactly how DEAH-box proteins induce remodeling events within RNPs; however, the evidence for DEAH-box stimulated RNA disassembly stands strong (Chen *et al*., 2001; Jankowsky *et al*., 2001; Kistler and Guthrie, 2001; Lardelli *et al*., 2010; Schwer, 2008). There are several proposed mechanisms for RNA helicase function, based on *in vitro* studies (Jankowsky *et al*., 2001), including a promising model where the helicase tracks processively along a single stranded RNA, displacing RNAs in its path. In the event that the helicase is less processive, RNP disruption may be a result of a force exerted by the helicase to disrupt an interaction between a protein and an RNA, and this single breakage may lead to a destabilization of a larger complex from the RNA (Jankowsky and Bowers, 2006). Furthermore, it has been demonstrated that DEAD-box binding of ATP can induce short RNA duplex unwinding, likely by inducing formation of a conformation of the DEAD-box protein favorable for strand separation (Chen *et al*., 2008). It is also possible that a conformational change upon ATP hydrolysis could lead to a physical interaction between protein and DExH/D-box factor serving to create the disruption. In any case, the consequences of the actions of DExH/D-box factors in splicing lead to conformational changes within the spliceosome, which permit forward progress in the splicing cycle.

## 1.3 *SACCHAROMYCES CEREVISIAE* AND SPLICING

In 1996, Goffeau *et al*. published the sequence of the first eukaryotic genome (Goffeau *et al*., 1996). This genome, belonging to the eukaryotic single cellular organism, *S. cerevisiae,* is comprised of roughly 6000 genes. Of these, 5885 genes are protein encoding, while 140 code for ribosomal RNAs, 40 for small nuclear RNAs, and 275 code for transfer RNAs (Goffeau *et al*., 1996). Interestingly, of the protein encoding genes, only ~5% contain introns, and the vast majority of these genes contain only a single intron. It has been shown that roughly one third of these introns are not essential and have only minor effects on cell growth (Parenteau *et al*., 2008). Meanwhile, the genomes of complex eukaryotes are mostly comprised of genes with at least one intron, and nine on average (Lander *et al*., 2001).

Despite the relatively low levels of genes carrying introns in *S. cerevisiae*, the introns that do exist are mostly found in genes that are highly expressed. The genes containing introns actually make up close to one third of all cellular transcripts (Ares *et al*., 1999), and it has been shown that introns actually increase the yield of both transcriptional and translational products (Juneau *et al*., 2006). Additionally, roughly 75% of all genes encoding ribosomal proteins contain introns, while introns are also found frequently in genes encoding secretion and meiosis factors (Juneau *et al*., 2006). It is interesting to note that ribosomal genes seem to have longer introns when compared to other classes of intron-containing genes (Spingola *et al*., 1999). These non-random characteristics of introns in *S. cerevisiae* build an intriguing, simplified model organism for the study of splicing.

The splice site sequences in yeast pre-mRNA are more highly conserved than those in higher eukaryotes (Meyer and Vilardell, 2009). The lack of defined splice sites in

complex eukaryotes is thought to be predominantly due to the need for alternative usage of splice sites in alternative splicing. With the conserved GUAUGU at the 5' splice site and the conserved branchpoint sequence of UACUAAC, yeast provide a defined system for studying the fundamentals of splicing, pertinent to all eukaryotic organisms.

Higher eukaryotes have the means by which a gene product can be alternatively spliced. The ability of higher eukaryotes to splice pre-mRNA along multiple pathways likely plays an important role in the evolution of their complexity; or perhaps it is evolution that allows them the adaptation to alternatively splice (Cooper *et al*., 2009). On the other hand, there are very few reports of alternative splice site selection in *S. cerevisiae* (Davis *et al*., 2000). Despite the substantial difference between *S. cerevisiae* and higher eukaryotes when it comes to alternative splicing, the splicing machinery and regulation for the splicing process include the same basic foundation (Meyer and Vilardell, 2009).

In addition to alternative splicing, the metazoan system has another layer of complexity, with the existence of a second "minor" spliceosome (Tarn and Steitz, 1997). This spliceosome uses similar principles in assembly and splicing mechanism; however, splicing of minor class introns operates on different platform. The minor spliceosome is composed U11 and U12, which are analogous to U1 and U2 in the major spliceosome, and U4ATAC and U6ATAC, which act in place of U4 and U6. The fifth snRNA involved in minor class splicing is the U5 snRNA, which is shared between the major and minor spliceosomes (Luo *et al*., 1999). Though yeast lacks this second spliceosome entirely, the splicing of both major and minor class introns relies on fundamentals, which have been shown and continue to be studied in *S. cerevisiae*.

Because it seems introns are the rule in higher eukaryotes rather than the exception, complex eukaryotic cells must accommodate by having high levels of the

machinery required for this process. Similarly, *S. cerevisiae* also proportionally accommodates the intronic levels within the cell with splicing machinery. This makes the isolation and characterization during biochemical studies of spliceosomal intermediates in yeast challenging, requiring nothing short of large-scale cellular growth.

The simplified splicing system of *S. cerevisiae* offers a number of other advantages as a model organism. First, yeast replicate rapidly, with wildtype doubling times of roughly 90 minutes (Botstein and Fink, 1988). Secondly, yeast readily undergo homologous recombination, making genetic manipulation, including insertions, deletions and integrative disruptions, possible and tractable (Botstein and Fink, 1988). The screening of either temperature sensitive mutant collections (Blanton *et al*., 1992; Lockhart and Rymond, 1994; Rosbash *et al*., 1981; Vijayraghavan *et al*., 1989) or cold sensitive mutant collections (Noble and Guthrie, 1996) for mutants who have splicing defects when subjected to non-permissive temperatures, have provided a means by which a substantial number of splicing factors have been discovered. Interestingly, many of the cold sensitive conditional splicing mutants prove very useful for biochemical analysis. This will be discussed in more detail in Chapter 2.

While *S. cerevisiae* is optimal for studying basic cellular processing events because of its simplicity, it is obviously a limited model, lacking the machinery that would likely develop the organism into a more complex eukaryote. Even so this organism provides an excellent system to study the foundation of what has proven to be a critical element of gene expression and regulation.

## 1.4 DISSERTATION OBJECTIVES

When my dissertation work began, the current understanding of the catalytic steps of the splicing cycle was limited to characterizations of particles that were reconstituted upon *in vitro* transcribed substrates. Although these particles provide a foundation from which the knowledge of splicing can be expanded, these particles do not necessarily represent authentic on-pathway intermediates of the splicing cycle because they are not necessarily functional intermediates. In addition, assembly upon a single transcript does not represent spliceosomal particles assembled upon the entire collection of yeast intron-containing transcripts. Because of the dynamic nature of the spliceosome, *in vivo* analyses are difficult, yet necessary from a biochemical standpoint. Accumulating homogenous spliceosomal particles at a level high enough to molecularly analyze is another challenging problem.

With my dissertation work, I intended to employ the power of yeast genetics to arrest spliceosomes *in vivo* and isolate the normally transiently occurring spliceosomal intermediates. My intention was to compositionally characterize *in vivo* purified particles. Compositional comparisons between particles made the characterization more interesting, as striking changes between spliceosomal intermediates became apparent. Because the particles were arrested at a non-permissive temperature some of the analysis allowed functional follow-up by incubating the purified particles under the proper conditions at permissive temperature.

This dissertation is divided into five chapters. Chapter one includes the discovery and history of the spliceosome and its constituents, and a detailed report of the splicing reaction in the context of the splicing cycle. This is followed by a review of the current understanding of the protein factors that are thought to affect the conformational

rearrangements within the spliceosome and permit the forward progress of the splicing cycle. Finally, this chapter provides a line of reasoning for using *S. cerevisiae* as a model organism to study splicing.

The second chapter outlines the first part of my dissertation project. In this chapter, the methods for arresting and isolating functional spliceosomes assembled *in vivo* will be introduced. Spliceosomes arrested before the first step of catalysis, before the second step of catalysis, and before the message is released from the spliceosome will be presented, and a compositional analysis and comparison of these isolated particles is provided.

Though the original goal of this project was to compare the components of catalytically arrested spliceosomes, the compositions of each spliceosome led to the hypothesis that a large complex of proteins, known as the SF3 complex, is released following one of the DExH/D-box remodeling events. Chapter 3 details functional testing of first-step arrested particles, and how a chase from an arrested particle through a single catalytic step provided insights a novel ATP-independent conformation change, in addition to the ATP-dependent release of the SF3 complex from the branchpoint sequence.

After the spliceosome has completed a round of splicing, it must disassemble into its snRNP components and reassemble. This recycling of snRNPs requires at least one ATP-independent rearrangement of U6 snRNA prior to its entry into the splicing cycle. Following the remodeling, U6 snRNA can base pair with U4 snRNA. These events are facilitated by an essential RNA binding protein, Prp24p. Chapter four details the genetic analysis of the essential nature of Prp24p and biochemical experiments on *in vivo* purified particles that led to a model where Prp24p may also be required for events following di-snRNP assembly in splicing cycle.

The major theme of my dissertation work has been isolating and characterizing dynamic spliceosomal particles using genetic strategies. By employing cold-sensitive mutant proteins, this work has demonstrated the purification and characterization of the major intermediates in the splicing cycle, in addition to previously uncharacterized assembly intermediates. These particles have been arrested *in vivo*, purified, compositionally defined, and functionally tested to reveal novel mechanistic information regarding pre-mRNA splicing and the recycling of snRNPs.

# Chapter 2: *In vivo* characterization of catalytic splicing intermediates

## 2.1 THE CATALYTIC STEPS OF THE SPLICING CYCLE

To fully appreciate and understand the spliceosome, it is crucial to define the mechanisms underlying its assembly, the authorization for initiating the two chemical steps and the proofreading of the splice sites. Although both pre-mRNA introns and group II introns possess a conserved 5' splice site, 3' splice site and branchpoint sequence, spliceosomal introns lack the sequences and structures that group II introns rely on to self-catalyze their intron removal. Instead, intron removal from pre-mRNAs depends on the regulated and coordinated efforts of over one hundred snRNP-associated and non-snRNP associated splicing factors. These factors contribute to the proper assembly of the substrate, the precise formation of the spliceosome and dynamic rearrangements within the catalytic core of the spliceosome. Among these factors are eight highly conserved RNA helicase-like proteins of the DExH/D-box family (Staley and Guthrie, 1998), and three of these, all DEAH factors, play a critical role during catalysis (Illustration 2.1). Prp2p is required in an ATP-dependent manner for the first chemical step (Chen and Lin, 1990; King and Beggs, 1990), while Prp16p uses ATP to permit the second chemical step (Schwer and Guthrie, 1991). Following catalysis the ATP- dependent function of Prp22p is required to release the spliced message (Company *et al*., 1991). Downstream of the conserved DEAH box domain, each of these proteins share a homologous C-terminal domain (Burgess *et al*., 1990; Chen and Lin, 1990; Vijayraghavan *et al*., 1989), which is important for interaction with the spliceosome (Edwalds-Gilbert *et al*., 2004) and for function (Schneider and Schwer, 2001). Although these factors have been assigned to particular events, the exact nature of the molecular rearrangements they initiate has remained enigmatic.

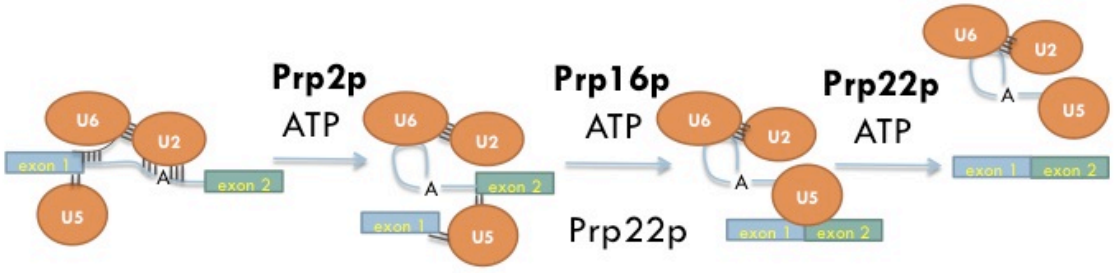


Illustration 2.1 DEAH-box proteins permit forward process during the catalytic steps of splicing.

Prp2p induces an ATP-dependent conformational rearrangement prior to the first step of splicing. Prp16p is required in an ATP-dependent manner to induce a rearrangement before the second chemical step. Prp22p acts in an ATP-independent manner in the second step of chemistry and is required to promote the release of the spliced message using its ATP-dependent function.

### 2.1.1 The first step of splicing

The first chemical step of splicing occurs when the 2'-OH of the intron branchpoint adenosine attacks the phosphodiester bond of the 5' splice site, releasing the 5' exon and forming a branched lariat intermediate RNA (Padgett *et al*., 1984; Ruskin *et al*., 1984). This reaction takes place in the active site of the spliceosome, and while studies focused on the second step of splicing have revealed active site components, the first-step factors and their role in catalysis have remained elusive.

Prp2p was implicated in splicing based on the finding that when mutant *prp2* heat inactivated extract was subjected to splicing conditions, pre-mRNA in the context of the 40S spliceosome accumulated (Lin *et al*., 1987). This result was followed up by complementing the extracts with wildtype extract containing elevated levels of Prp2p, and by adding ATP. This effectively chased the pre-mRNA into the products of the first reaction (Lin *et al*., 1987), implying a function for Prp2p in the first step. Later, the specific ATP requirement in the first step was demonstrated to be that of Prp2p, as this factor showed an RNA-dependent ATPase function specifically associated with the first step of splicing (Kim and Lin, 1993; Kim *et al*., 1992).

Further studies have shown that Prp2p only transiently interacts with the spliceosome, binding in the absence of ATP, and exiting upon ATP hydrolysis (Kim and Lin, 1993; King and Beggs, 1990). Upon ATP hydrolysis, Prp2p is thought to induce a conformational change in the spliceosome, which permits the second step of splicing (Kim and Lin, 1993; Kim and Lin, 1996). Evidence suggests that ATP hydrolysis by Prp2p is tightly linked to the first step of splicing; however, uncoupling these events presents a challenge.

### 2.1.2 The second step of splicing

The second step of the splicing reaction occurs when the 3'-OH of the 5' exon attacks the phosphodiester bond at the 3' splice site, joining the exons and liberating the intron lariat RNA. A number of protein factors, including Prp18p, Slu7p, Prp17p, and Prp22p have been characterized to participate in the second step in an ATP-independent manner. On the other hand, ATP hydrolysis by Prp16p is required prior to the second step of chemistry.

It has been proposed that the spliceosome resulting from Prp16p action prepares an active site for the second step chemistry (Umen and Guthrie, 1995c). Like Prp2p, ATP hydrolysis by Prp16p is an RNA-dependent reaction; however, purified Prp16p does not have RNA helicase capabilities (Schwer and Guthrie, 1991), suggesting its role is not in unwinding RNA. Instead, Prp16p is thought to carry out a conformational change that results in the correct positioning of the substrate for second step catalysis. Upon interaction with the spliceosome, Prp16p has been found to interact with the intron 3' splice site by crosslinking studies. Following hydrolysis Prp16p leaves the spliceosome (Schwer and Guthrie, 1991), showing a transient interaction prior to the second step analogous to the transient interaction of Prp2p before the first step of catalysis. This Prp16p-induced conformational rearrangement results in protection of the 3' splice site from oligonucleotide-directed RNase H cleavage (Schwer and Guthrie, 1992), implying that an active site may be forming around the second step nucleophile. Following this rearrangement, the second step factors Slu7p and Prp8p have been shown to crosslink at the 3' splice site, and to act just before the chemical reaction (Umen and Guthrie, 1995b).

The action of Prp16p is required for the second step of catalysis to proceed; however, ATPase-deficient mutant alleles of *prp16* are suppressed by branchpoint mutants. This is thought to occur not by kinetically increasing the rate of the second step,

but instead by inhibiting the degradation of the mutant lariats (Burgess and Guthrie, 1993), thus assigning a proofreading mechanism to Prp16p in the second step of splicing.

### 2.1.3 Message release

An ATP-independent role has been assigned to Prp22p in the second catalytic step; however, this role is only essential for introns containing large distances between the branchpoint and the 3' splice site (Schwer and Gross, 1998). On the other hand, Prp22p is essential for the ATP-dependent dissociation of the mRNA from the spliceosome (Company *et al*., 1991). In mRNA displacement, Prp22p has been proposed to bind downstream of the 3' splice site and act to disrupt the mRNA/U5 snRNP contacts using 3' →5' helicase activity (Schwer, 2008).

### 2.1.4 Characterizations of dynamic intermediates of splicing

The dynamic nature of the spliceosome, especially during the catalytic stages of the splicing cycle makes studying splicing intermediates a major challenge. Previous reports have focused their efforts upon the analysis of spliceosomes assembled from extracts on *in vitro* transcribed substrates. Experimental approaches include immunodepleting an essential factor (Chiu *et al*., 2009; Liu *et al*., 2007) or using transcripts with mutations in the conserved splice sites or branchpoint sequences (Bessonov *et al*., 2008; Fabrizio *et al*., 2009). A recent study, published while this dissertation work was under review, inactivated splicing in extracts made from temperature sensitive *prp2* strain (Warkocki *et al*., 2009). Though these strategies provide a foundation for analyzing splicing intermediates, particles assembled from extracts do not necessarily represent authentic, on-pathway, functional intermediates from the

splicing cycle. Additionally, the characterization of these intermediates is limited to complexes assembled upon a single pre-mRNA transcript.

Though the dynamic process of splicing creates a hurdle for understanding the role of the spliceosome in splicing, it is also critical to understand how the splicing machinery acts in concert with pre-mRNA to catalyze intron removal from spliceosomes assembled during normal cellular activity. This chapter described a genetic strategy, which allows for biochemical purification of functional, synchronized spliceosomes from yeast, assembled *in vivo*. Spliceosomes were genetically arrested at three critical checkpoints, namely, spliceosomes arrested before the first and second steps of splicing and spliceosomes arrested post-catalysis, to enrich for on-pathway, functional complexes. This chapter describes the characterized components of these spliceosomes, accounting for the snRNA, protein and pre-mRNA content. These compositional studies provide the first study of *in vivo* assembled spliceosomes during the catalytic stages of splicing (Lardelli *et al*., 2010).

## 2.2 MATERIALS AND METHODS

### 2.2.1 Tagging vectors

| Vector | Purpose | Selectable marker |
|---|---|---|
| pJPS1162 | TAP tagging | *KANr (Kanamycin resistance)* |
| pBS1539 | TAP tagging | *URA* |

Table 2.1 Plasmids used generating TAP tags in Chapter 2.

### 2.2.2 Yeast strains

| Strain | Mating type | Genotype |
|---|---|---|
| SS1347 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*PRP2-TAP::URA3* |
| SS1348 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*prp2G551N-TAP::URA3* |
| SS492 | a | *hisΔ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, PRP16-TAP::TRP1* |
| SS1316 | a | *ade2, his3, leu2, lys2, trp1, ura3, prp16::LYS* pSE358-*prp16-302-TAP::URA3* |
| SS1311 | α | *gal-, mal-, his4-619, prp16R456K-TAP::KAN* (parent DBY4406) |
| SS1330 | α | *gal-, mal-, his4-619, prp22D613Q-TAP::KAN* (parent DBY4172) |
| SS1332 | α | *gal-, mal-, his4-619, prp22-G394R, D613Q-TAP::KAN* (DBY4423) |
| SS481 | a | *PRP22-TAP::URA3* |

Table 2.2 Yeast strains used in Chapter 2.

### 2.2.3 Oligos

| Oligo name | Sequence |
|---|---|
| PRP16Gould TAPA | 5'ACGGCAAAGAAAATTCAATGAAACCTTTCAAAAGAAGGAAGCCTTTTTTTATGAAGCGACGATGGAAAAAG3' |
| PRP16Gould TAPB | 5'CATATAACTATATAATAACATATATGAATATTTTGCCTATTAGCACGCTCTACGACTCACTATAGGGA3' |
| PRP16TAPA | 5'ACGGCAAAGAAAATTCAATGAAACCTTTCAAAAGAAGGAAGCCTTTTTTTTCCATGGAAAAGAGAAG3' |
| PRP16TAPB | 5'TGCATATAACTATATAATAACATATATGAATATTTTGCCTATTAGCACGCTACGACTCACTATAGGG 3' |
| PRP2TAPA | 5'CACAAATCTTTAAAGATTTAATTGACGATAAAACAAATAGGGGGAGGCGGTCCATGGAAAAGAGAAG3' |
| PRP2TAPB | 5'TGCATATAGAATGGAGCCTGCGTTTCTAGCAATACACATACACCTGTCAATACGACTCACTATAGGG3' |
| PRP22Gould TAPA | 5'GACTAAGCTCAATAAGGCAGTCAAGGGAAAGGGCATTAGGTATCAAGAGGATGAAGCGACGATGGAAAAAG 3' |
| PRP22Gould TAPB | 5'TTAAAAAATTAAATATAGGTCTATAAAACTCGATAATTATAATGCATAAATACGACTCACTATAGGGA3' |
| U1 probe | 5'GAATGGAAACGTCAGCAAACAC3' |
| U2 probe | 5'AAGAACAGATACTACACTTGA3' |
| U4 probe | 5'ACCATGAGGAGACGGTCTGG3' |
| U5 probe | 5'ATGTTCGTTATAAGTTCTATAGGC'3 |
| U6 probe | 5'AGGGGAACTGCTGATC3' |
| RPP1B ex1SS5 | 5'ATTGACTATCACCAAGGCCGC3' |
| RPP1B ex2-3 | 5'GTGACGCTGCTGCTGAAGAAGAAA3' |
| YFL039C_1_p_F | 5'AGGGGCTTGAAATTTGGAAAAA 3' |
| YFL039C_1_L_R | 5'GCAAGCGCTAGAACATACATAGTACA3' |
| YFL039C_1_I_F | 5'TTGCTTCATTCTTTTTGTTGCT3' |
| YFL039C_1_I_R | 5'GCAAAACCGGCTTTACACAT3' |
| PRP22seq1 | 5'TATGAAGGTAAAGTGAGAAAC3' |
| PRP22seq2 | 5'ACAACAACGAGACGAGACGGATG3' |
| PRP22seq3 | 5'CCTCGTAGGGTTGCCGCTGTATC3' |
| PRP22seq4 | 5'AGGTGACAGCATCGGCGAATTAC 3 |
| PRP22seq5 | 5'CACGCTCTTTATTGTCATCTGTTG3' |
| PRP16seq1 | 5'GGTGCCTGAACCTCTGTCTGAATT 3' |
| PRP16seq2 | 5'ATACTAAAGAGCAACTGCCTG 3' |
| PRP16seq3 | 5'TATACTTCCAACCCTGTTCAAGAC3' |
| PRP16seq4 | 5'TCCAAAGAACGAATTTGTCAAAC3' |
| PRP16seq5 | 5'ATGTATTTGTTCAGGATTTGCTC3' |
| PRP2PCRA | 5'TACTAGCCCATACAACCTGATGAC3' |
| PRP2seq1 | 5'CAGACAACGCATGTGGGAAG3' |
| PRP2seq2 | 5'CATGAACGTACTCTAGCCACAG3' |
| PRP2seq3 | 5'CTCCACTGGAATGACCCAAC3' |
| PRP2seq4 | 5'GTGTGTTAAGCGAAGTGGAAAG3' |
| PRP2PCRB | 5'CGCATACCACATCTTCAAACG3' |

Table 2.3 Oligos used for PCR generation of tags, Northern blotting, RT-PCR, and sequencing in Chapter 2.

### 2.2.4 Creation of yeast strains

Yeast strains listed in section 2.2.2 were created by through use of a PCR-generated fragment with arms of homology to the region in which the tag was to be knocked in, followed by a selectable marker, *URA3* for TAP tags generated from pBS1539, or *KANr* (*Kanamycin resistance)* for TAP fragments generated from pJPS1162. Fragments were transformed in using the lithium acetate transformation method (Gietz and Woods, 2002). Colonies that grew on synthetic media lacking *URA* or on rich media containing Geneticin (Gibco BRL Life Technologies, Inc.), were screened for presence of a functional TAP tag by western blotting. In brief, whole cell extract was subjected to SDS-PAGE, transferred to nitrocellulose (Bio-Rad) overnight at 4°C (in 20% methanol, 0.001% SDS, 386 mM glycine, 48 mM Tris base), blocked with BLOTTO (PBST + 5% SDS), washed twice in PBST for 5 minutes each, incubated with a 1:1000 dilution of the peroxidase anti-peroxidase antibody (Rockland) for 1 hour, washed with PBST 4 times for 15 minutes each, subjected to Chemiluminescent detection (Perkin Elmer), and exposed to film.

### 2.2.5 Mapping helicase mutations

To map the mutations in the *prp16* and *prp22* genes in the DBY strains, the entire open reading frame was PCR-amplified from the respective strains. The *prp16* mutation contained in DBY4406 is Arg 456 to Lys, located just downstream from motif Ib. The mutation in *prp22* contained in DBY4172 is Asp 613 to Gln, a charge reversal located just after the DEAH-box (motif II). The mutation in *prp22* contained in DBY4423 is Gly 394 to Arg, as well as the same Asp 613 to Gln also found in DBY4172. The

characterized *prp2* mutation Gly 551 to Asn is accompanied by an additional previously uncharacterized mutation: Asp 547 to Glu, which suppresses the dominant negative phenotype reported by Lin and colleagues (Edwalds-Gilbert *et al*., 2000).

### 2.2.6 Cell growth and temperature shifts

Wildtype strains were grown in YPD at 31°C to OD ($A_{600}$) = 1.5 before harvesting. Mutant strains were grown at PT (*prp2G551N* strains at 37°C; *prp16* and *prp22* strains at 31°C) to OD ($A_{600}$) = 1.0 and then shifted to NPT (*prp2G551N* strains at 25°C; *prp16* and *prp22* strains at 16°C) for 45 minutes. Cells were harvested by centrifugation, washed in cold water, washed in buffer A (10mM HEPES, pH 7.9, 200 mM KCl, 1.5 mM $MgCl_2$, 8% glycerol), and resuspended in one volume (w:v) buffer A before freezing dropwise in liquid nitrogen.

### 2.2.7 Isolation of arrested spliceosomes

The Tandem Affinity Purification (TAP) tag provides a method for two-step purifications of complexes assembled *in vivo* (Rigaut *et al*., 1999). The TAP tag combines two affinity tags, separated by a Tobacco Etch Virus (TEV) cleavage site (Dougherty *et al*., 1989)(see Illustration 2.2). With a Protein-A domain at the C-terminus, the complex of interest is captured on IgG antibody-conjugated beads. Following washes, the Protein A tag can be cleaved from the complex, and left on the column through of the TEV protease cleavage site, which lies in between the protein-A tag and the Calmodulin binding domain. The complex can then be affinity purified to a further extent by using the Calmodulin resin, and eluting with EGTA (Niggli *et al*., 1979). The affinity steps in this dissertation employ the use of only the Protein-A tag and the TEV cleavage site of the

TAP tag. As described below, glycerol gradient sedimentation serves a second and size separating purification step following the initial affinity purification, which keeps the *in vivo* assembled particle in tact.

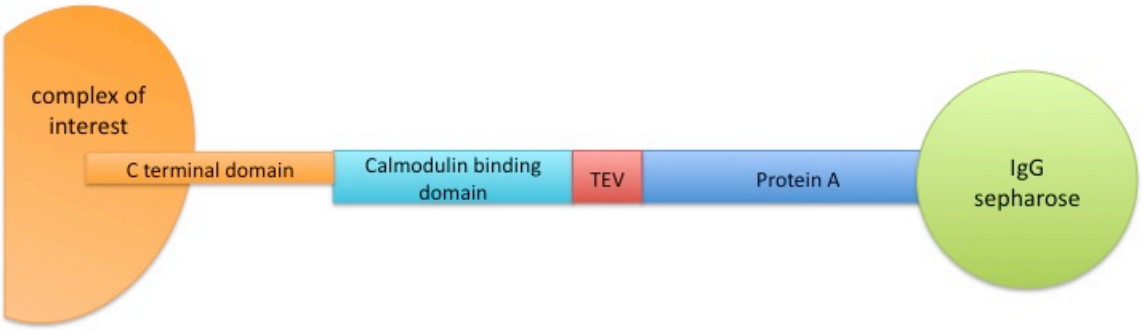


Illustration 2.2 The TAP tag can be used for two-step affinity purification of native *in vivo* assembled complexes.

An IgG affinity column captures the Protein-A tag at the C-terminus of a protein contained within the complex of interest. The Protein-A tag can be cleaved by TEV protease, leaving the complex of interest with a C-terminal Calmodulin binding domain. This can be used in a secondary purification step.

Frozen cells were milled using a Retsch mixer mill (Mayas *et al*., 2006) with 50 mL canisters immersed in liquid nitrogen. Briefly, cells were disrupted by shaking at 10 Hz for 5 cycles of 3 minutes each. ATP was depleted by addition of glucose to 2 mM. Whole cell extract was prepared as described (Stevens and Abelson, 2002). Briefly, the pulverate was thawed, and allowed to extract for 30 minutes by gentle rotation at 4°C in the presence of 5 μM Leupeptin, 3 μM Pepstatin, 0.4 mM PMSF, and 1 mM DTT. Extract was cleared by centrifugation at 30,000 x g for 30 minutes at 4°C, and then 100,000 x g for 1 hour at 4°C. The middle third of the cleared extract was gently removed following the second centrifugation step, and allowed to dialyze in Buffer D-Lite (containing 8% glycerol, 150 mM KCl, 1.5 mM $MgCl_2$, 20 mM HEPES, pH 7.9) for 2 x 1 hour. Extract (~40 mL) was incubated with IgG sepharose (400 μL) for 1 hour at 4°C with rotation to allow binding of the TAP tag to IgG. The resin was washed with 40 mL of IPP150 (10 mM Tris, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1.5 mM $MgCl_2$, 8% glycerol, 1 mM DTT), and then resuspended in IPP150 (1.2 mL). Bound material was incubated for 1 hour at 16°C with TEV protease to cleave the TAP tag and RNase Inhibitor (Invitrogen). The eluted material was layered onto 10-30% glycerol gradients (made with 20 mM HEPES, pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$, 0.2% NP-40, 1 mM DTT, 0.2 mM PMSF), and the gradients were centrifuged for 10 hours at 29,000 RPM in an SW41 rotor at 4°C. Fractions were phenol/chloroform extracted, and protein and RNA were precipitated (Stevens and Abelson, 2002).

### 2.2.8 Northern blot analysis

RNA was precipitated by adding 2.5 volumes of ethanol, and incubating overnight at -20°C. RNA was pelleted by centrifugation and washed with 70% ethanol. Following drying the pellet, the RNA was resuspended in formamide loading buffer (80% formamide, 10mM EDTA pH 8.0, xylene cyanol, bromophenol blue), heated for 10 minutes at 65°C and then separated by electrophoresis on a 7% 19:1 polyacrylamide gel, containing 8M urea and 1x TBE, for 1 hour and 45 minutes at 25 mA/gel. Gels were transferred to nylon membrane (Whatman) overnight at 0.5 A in 25 mM $Na_2HPO_4$, pH 7.0 buffer at 4°C. The RNA was crosslinked to the membrane by exposing the membrane to UV (254 nm) light for 1 minute, prior to pre-hybridization incubation at 42°C in Church buffer (500 mM $Na_2HPO_4$ pH 7.0, 7% SDS, 1 mM EDTA)(Church and Gilbert, 1984). Hybridization with $\gamma$ $^{32}$P radiolabeled oligonucleotides was performed overnight at 42°C. The membrane was then washed briefly in 6x SSC, 2 times with 2x SSC + 0.5% SDS for 15 minutes each at 42°C, and once more briefly with 6x SSC. The membranes were then exposed to a phosphorimaging screen (Bio-Rad), and developed in a phosphorimager (GE Healthcare Typhoon).

### 2.2.9 RNA quantitiation and normalization

Quantitation of snRNAs was performed by phosphorimaging analysis using the Quantity One suite of software (Bio-Rad). Values for each band within the Northern blot were quantitated using identically sized analysis frames. Background levels were subtracted from each of the values using an identically sized analysis frame in a region of the Northern blot that contained no signal. Background-subtracted RNA signals from

each glycerol gradient fraction were normalized against the lane with the most intense signal for that Northern blot, which gave a value of one (1) for the most abundant fraction.

### 2.2.10 RT-PCR

Peak fractions from the glycerol gradients were treated with DNase I (Promega) and RNase inhibitor (Invitrogen) for 15 minutes at 37°C. RNA was phenol/chloroform extracted, precipitated, and reverse transcribed with AMV reverse transcriptase (Promega) using an oligonucleotide primer specific to the *RPP1B* ex2-3 oligo for full length transcripts or YFL039C_1_L_R (*ACT1*) for lariat or YFL039C_1_I_R (*ACT1*) for the 3' splice site (J. Beggs and D. Barrass, personal communication). PCR on cDNA was performed using Pfu DNA polymerase using the above primers with *RPP1B* ex1SS5, YFL039C_1_p_F or YFL039C_1_I_F, respectively, to detect the species of interest.

### 2.2.11 Mass spectrometry analysis

Mass spectrometry analysis of proteins isolated from glycerol gradient fractions corresponding to the spliceosome peaks was carried out by James Thompson in the Yates laboratory at The Scripps Research Institute. A detailed description of the materials and methods can be found in (Lardelli *et al*., 2010) and is provided in Appendix I.

In brief, lyophilized protein samples were re-solubilized, reduced with TCEP and alkylated with iodoacetamide prior to tryptic digests. Samples were desalted and separated by a modified six-step HPLC protocol with a two-dimensional conjugated C18/strong cation exchange column. Peptides were eluted and electrosprayed into a two-dimensional ion trap mass spectrometer (ThermoFinnigan). Eight data-dependent MS/MS

spectra were taken and analyzed using SEQUEST algorithm (Eng *et al*., 1994) against the *Saccharomyces cerevisiae* genome database (SGD).

Parameters for inclusion in the data presented in Tables 2.4, 2.5 and 2.6 include: sequence confirmation of >5% of protein and absence from the mock purification MS/MS data. Mock purification was performed identically to that of the experimental samples from a nontagged strain of yeast. Polypeptides identified in this mock analysis are presented in Appendix II, and contaminants identified in the *prp2, prp16* and *prp22* particles are presented in Appendix III.

## 2.3 RESULTS

### 2.3.1 Arresting first step-blocked spliceosomes

Prp2p acts almost immediately upstream of the first step of pre-mRNA splicing; therefore, we isolated spliceosomes from *S. cerevisiae* arrested prior to Prp2p function. Prp2p-associated material was affinity purified from cells containing a C-terminal TAP-tag (Puig *et al*., 2001; Rigaut *et al*., 1999) at the *PRP2* locus in a wild-type strain. Prp2p-associated complexes were layered onto a glycerol velocity gradient suitable for spliceosome sedimentation (Stevens *et al*., 2002). Northern blot analysis of snRNAs in the even fractions from this gradient shows that Prp2p does not stably interact with significant levels of endogenous spliceosomal snRNAs, as represented by the lack of U2, U5 and U6 in the glycerol gradient (Figure 2.1A). As Prp2p is a transiently interacting factor (King and Beggs, 1990), we did not expect this strategy to yield spliceosomes, but instead to serve as a control.

To accumulate spliceosomes which have Prp2p stably associated we used a cold sensitive (cs) *prp2* that interacts with spliceosome. A yeast strain containing the severely cold sensitive G551N mutation of Prp2p (Edwalds-Gilbert *et al*., 2000), which is defective ATP hydrolysis and release from the spliceosome, was C-terminally TAP-tagged at the *prp2* locus. Cells were shifted to the NPT of 25°C prior to harvesting, trapping endogenous pre-mRNAs in a first step-arrested spliceosome. ATP depleted, affinity-purified prp2G551Np-associated material was layered onto a glycerol velocity gradient for spliceosome isolation. A significant peak of U2, U5 and U6 snRNAs at ~40S is seen in these gradients (Figure 2.1B) corresponding to size of the yeast spliceosome (Brody and Abelson, 1985).

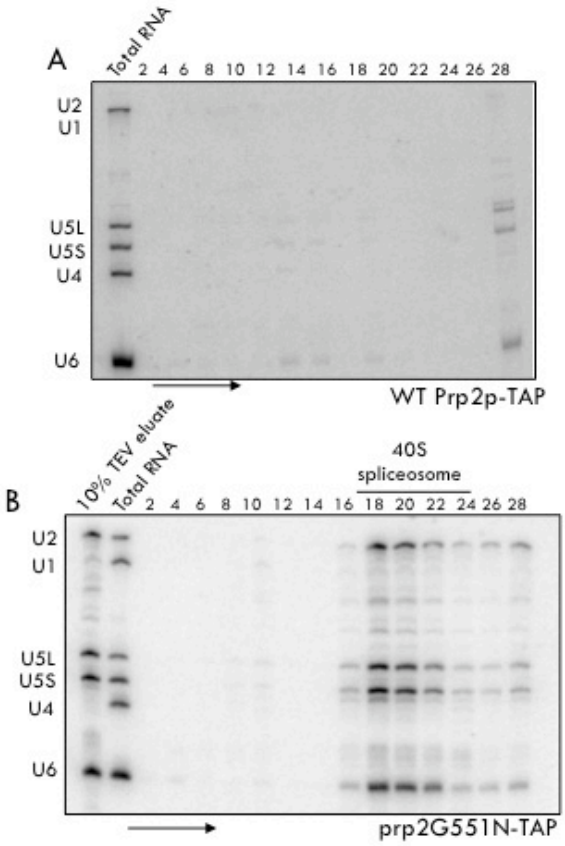


Figure 2.1 prp2G551Np mutant yeast accumulate snRNAs in the ~40S region of the glycerol gradient.

A. Prp2p-TAP affinity purified, glycerol gradient-separated material contains no significant accumulation of snRNAs in the ~40S region of the glycerol gradient, after Northern blotting for all five snRNAs, indicating the absence of associated spliceosome.

B. prp2G551Np-TAP-associated material was Northern blotted for spliceosomal snRNAs. Splicing complexes, as indicated by the presence of U2, U5 and U6 are seen in the ~40S region of the glycerol gradient. The arrow indicates the direction of sedimentation, and positions of U1, U2, U4, U5 and U6 snRNAs are shown.

### 2.3.2 Determination of pre-mRNA splicing status within *prp2*-arrested particle

RNA from the peak fraction of the 40S spliceosome sized particle was subjected to RT-PCR with primers against an intron-containing transcript, *RPP1B*. Figure 2.2 shows that the intron-containing transcript, *RPP1B* pre-mRNA is fully unspliced at this point in the splicing cycle, implying that the purified spliceosome is indeed arrested at the first chemical step. Functional characterization of this transcript included incubation at permissive temperature (PT) or 37°C in the presence of ATP for 20 minutes. Interestingly, the data indicates that *RPP1B* can be functionally chased through both steps of splicing, indicating one of two possibilities: either the proteins present are sufficient to proceed through both steps of splicing, or there is a minute presence of second step factors that permit both steps of splicing.

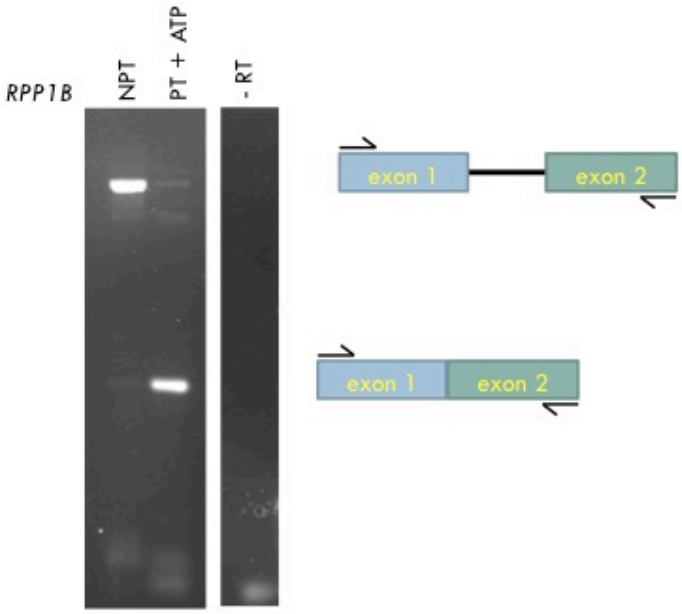


Figure 2.2 *prp2*-arrested spliceosome contain pre-mRNA, which can be chased to a spliced message as determined by RT-PCR.

RT-PCR of an intron-containing transcript was used to assess the splicing status of the intron within the *prp2*-arrested particle. At NPT, *RPP1B* is unspliced. After incubating the particle at PT in the presence of ATP for 20 minutes, the pre-mRNA is chased to a fully spliced message. A no-RT control shows no pre-mRNA or mRNA amplification.

### 2.3.3 Compositional characterization of first step-arrested spliceosomes

To compositionally define the first step-arrested spliceosome, we performed MudPIT mass spectrometry analysis of the associated polypeptides (Washburn *et al*., 2001) from the fractions of the glycerol gradient corresponding to the spliceosome sized particle. Table 2.4 lists the components of the first step-arrested spliceosomes. Notably, all but three tri-snRNP components are absent in the *prp2*-arrested spliceosome. Additionally, the first step-arrested spliceosome contains all of the known yeast SF3 components (Dziembowski *et al*., 2004) and the U2 core proteins Lea1p and Msl1p (U2A' and U2B", respectively in human) (Caspary and Seraphin, 1998). Also present in the first step blocked spliceosome are the Sm core proteins of the U2 and U5 snRNPs (Seraphin, 1995). It is also worth mentioning that the U6-specific LSm proteins (Achsel *et al*., 1999) are absent, which is consistent with their destabilization during spliceosome activation (Chan *et al*., 2003). The entire Prp19p complex (NTC) is also present, as expected from the reported role of the NTC in stabilization of U5 and U6 in the events leading up to the first step of splicing (Tarn *et al*., 1994). Consistent with a recent report, the second step factor, Prp17p, is also detected prior to the first step (Sapra *et al*., 2008). Two pre-mRNA binding proteins (Sto1p and Cbc2p) comprising the cap-binding complex (Colot *et al*., 1996; Fortes *et al*., 1999; Lewis *et al*., 1996) are detected. As shown in Table 2.4, a number of other spliceosome-associated proteins are also present in the protein sample representing the first step-arrested spliceosomes.

| snRNP or complex association | Protein | # of peptides identified/percentage polypeptide |
|---|---|---|
| U2 snRNP core | Lea1p | 7/27.7% |
| | Msl1p | 4/29.7% |
| SF3b | Rse1p | 20/13.7% |
| | Hsh155p | 20/18.4% |
| | Cus1p | 27/39.7% |
| | Hsh49p | 4/24.4% |
| | Rds3p | 2/21.3% |
| | Ysf3p | 2/25.9% |
| SF3a | Prp21p | 13/29.3% |
| | Prp11p | 9/30.1% |
| | Prp9p | 38/47.7% |
| U4/U6•U5 | Prp8p | 60/20.3% |
| | Brr2p | 56/21.7% |
| | Snu114p | 27/26.0% |
| U2, U5 snRNPs | Sm core | ✓ |
| NTC | Prp19p | 28/40.8% |
| | Prp46p | 16/28.6% |
| | Snt309p | 13/41.7% |
| | Cef1p | 36/53.7% |
| | Syf1p | 13/15.9% |
| | Clf1p | 14/17.6% |
| | Ntc20p | 5/24.3% |
| | Cwc2p | 14/35.4% |
| | Isy1p | 6/31.5% |
| | Syf2p | 11/49.3% |
| 2nd step factors | Prp17p | 15/28.4% |
| Helicase/related | Prp2p | 59/41.9% |
| | Spp2p | 13/50.3% |
| RES complex | Bud13p | 14/34.6% |
| | Pml1p | 3/19.6% |
| | Ist3p | 3/26.4% |
| mRNA binding | Sto1p | 8/7.8% |
| | Cbc2p | 5/19.7% |
| Spliceosome associated | Bud31p | 7/17.2% |
| | Prp45p | 21/33.5% |
| | Cwc22p | 10/13.2% |
| | Cwc24p | 15/44.8% |
| | Cwc27p | 12/33.6% |
| | Cwc15p | 6/40.6% |
| | Cwc21p | 5/28.9% |
| | Ecm2p | 7/22.8% |

Table 2.4 Mass spectrometry analysis of proteins associated with the first step-arrested spliceosome from peak fraction of Figure 2.1B.

### 2.3.4 Second step-arrested spliceosome characterization

After the first step of splicing, Prp16p is required to proceed through the second step. Prp16p was TAP-tagged in three strains: BY4741 (wild-type *PRP16*), *prp16-302* (*prp16R456K, G691R*) (Madhani and Guthrie, 1994), and the allele of *prp16* in strain DBY4406 (*prp16R456K*; (Noble and Guthrie, 1996). Consistent with its transient interaction with the spliceosome and similar to wildtype Prp2p-TAP (Figure 2.1A), affinity-purified, wildtype Prp16p-TAP is not associated with significant levels of spliceosomal snRNAs in any region of the glycerol gradient (Figure 2.3A). Spliceosomes purified from both cold sensitive *prp16* strains grown at NPT are associated with U2, U5 and U6 snRNAs in the ~40S region of the glycerol gradient (Figure 2.3B).

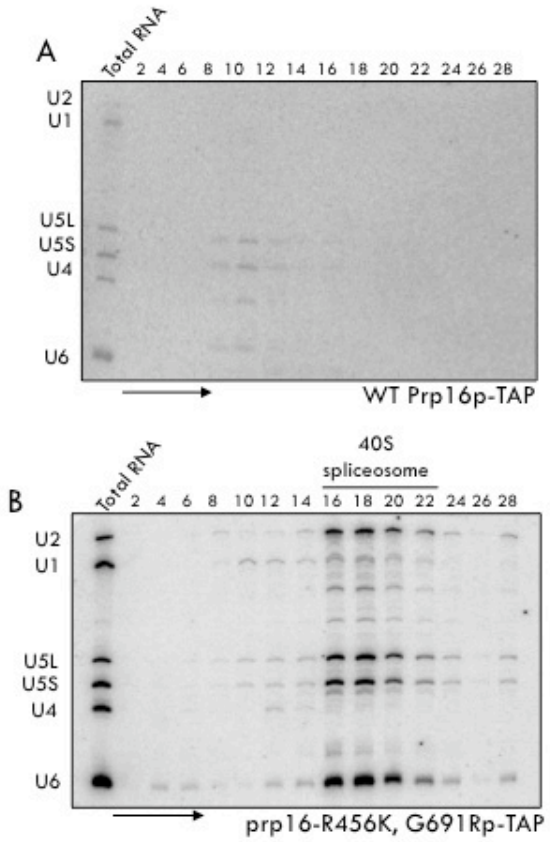


Figure 2.3 Cold-sensitive *prp16* alleles, accumulate spliceosome in the ~40S region of the glycerol gradient.

A. Prp16p-TAP affinity-purified, size-separated material contains no significant accumulation of snRNAs, after Northern blotting for all five snRNAs across the glycerol gradient.

B. prp16-R456K, G691Rp-TAP-associated, size separated material was Northern blotted for spliceosomal snRNAs. Splicing complexes, as indicated by the presence of U2, U5 and U6, are seen in the ~40S region of the glycerol gradient. The arrow indicates the direction of sedimentation, and positions of U1, U2, U4, U5 and U6 snRNAs are shown.

### 2.3.5 Pre-mRNA characterization of second step-arrested particle

RT-PCR analysis of the mRNA/pre-mRNA/lariat contained in the peak fractions from the second step-arrested particle was performed. Using reverse transcriptase primers that anneal to the 3' exon of *RPP1B,* the 3' exon of *ACT1*, or the branched sequence of the lariat of *ACT1*, primer extension was conducted. The RT products were then subjected to PCR using the following primers pairs: a 5' primer that anneals to the 5' exon of *RPP1B* and the 3' primer annealing to the 3' exon of *RPP1B* to detect full length substrate; a 5' primer that anneals to the intron just downstream of the branch sequence of *ACT1* and 3' primer that anneals to the 3' exon of *ACT1* to detect 3' splice site; or a 5' primer within the lariat of *ACT1* and the 3' primer annealing to the branched sequence of *ACT1* lariat to detect the lariat structure. As shown in Figure 2.4, the particle contains no pre-mRNA, but does contain the 3' splice site junction and the lariat RNA, indicative of the products of the first step of splicing. This confirms the isolation of a C-complex.

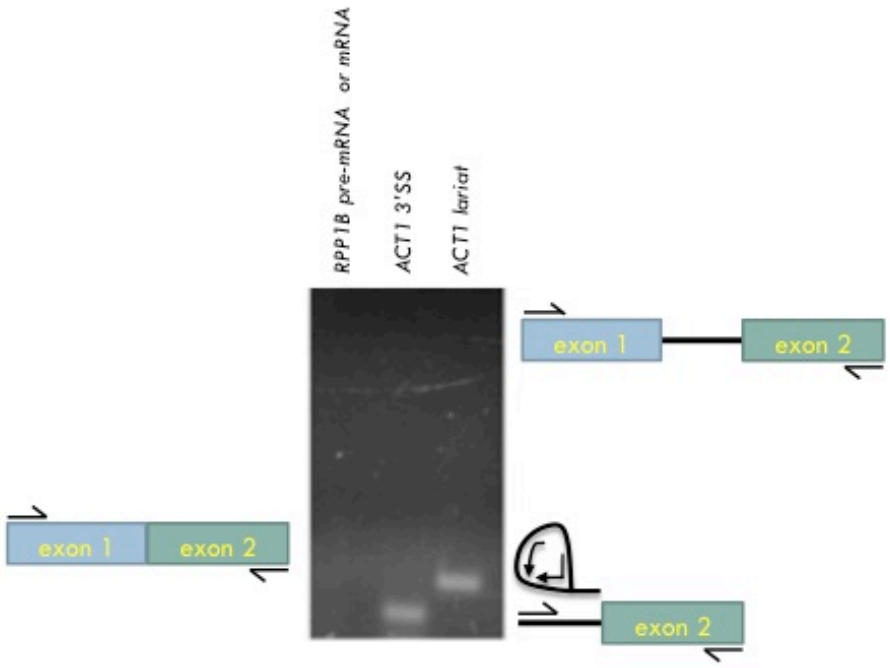


Figure 2.4 Purified *prp16*-arrested particles contain lariat intermediate as determined by RT-PCR.

RT-PCR of intron-containing transcripts was used to assess the splicing status of the intron within the *prp16*-arrested particle at NPT. The first set of primers anneal to the exons of *RPP1B*, and the lack of a band indicates there is neither pre-mRNA nor spliced message. Primers annealing to the *ACT1* intron and the 3' exon reveal the presence of a 3' splice site, which in the absence of pre-mRNA suggests the presence of lariat intermediate. Finally, primers that are designed to amplify a lariat structure of the *ACT1* intron show the presence of a lariat. Together, these finding confirm that the first step of splicing, but not the second step, has been completed in the arrested particles.

### 2.3.6 Compositional definition of the second step-arrested particle

Protein fractions corresponding to the second step-blocked particles were analyzed by MudPIT mass spectrometry analysis. Table 2.5 lists the protein components of the spliceosome prior to the second chemical step of splicing (Figure 2.3B). Unlike the first step step-arrested spliceosome (Figure 2.1B), the second step-arrested spliceosome is devoid of all SF3 components. This independently supports that both SF3a and SF3b are displaced from the spliceosome between these steps. Other differences include the dissociation of the Retention and Splicing complex (RES) complex (Dziembowski *et al*., 2004). Additionally, we note the presence of the proposed first step factors, Yju2p and Cwc25p, only after the first step of splicing has taken place, which is consistent with their assigned roles in the first step of chemistry after Prp2p ATP hydrolysis (Chiu *et al*., 2009; Liu *et al*., 2007). The disassembly factor, Prp43p (Martin *et al*., 2002), is added prior to the second chemical step.

| snRNP or complex association | Protein | # of peptides identified/ % polypeptide |
|---|---|---|
| U2 snRNP core | Lea1p | 16/53.4% |
| | Msl1p | 3/36.9% |
| U4/U6•U5 | Prp8p | 45/20.8% |
| | Brr2p | 55/24.6% |
| | Snu114p | 26/31.8% |
| U2, U5 snRNPs | Sm core | ✓ |
| NTC | Prp19p | 43/66.8% |
| | Prp46p | 25/47.7% |
| | Snt309p | 5/42.3% |
| | Cef1p | 14/26.4% |
| | Syf1p | 16/19.8% |
| | Clf1p | 15.27.5% |
| | Ntc20p | 6/20.7% |
| | Cwc2p | 20/44.0% |
| | Isy1p | 6/30.6% |
| | Syf2p | 11/35.3% |
| 2nd step factors | Prp17p | 7/20.4% |
| Helicase/related | Prp16p | 74/47.0% |
| | Prp22p | 6/7.1% |
| | Prp43p | 2/3.1% |
| THO/TREX | Yra1p | 7/27.4% |
| | Hpr1p | 2/4.0% |
| | Thp2p | 2/10.3% |
| | Mft1p | 3/6.9% |
| mRNA binding | Sto1p | 7/7.3% |
| | Cbc2p | 9/27.9% |
| Spliceosome associated | Yju2p | 8/31.3% |
| | Bud31p | 4/28.0% |
| | Prp45p | 21/43.0% |
| | Cwc22p | 12/24.4% |
| | Cwc25p | 12/50.3% |
| | Cwc15p | 8.41.7% |
| | Cwc21p | 6/31.1% |
| | Ecm2p | 17/46.4% |
| Novel | Ubp3p | 4/6.1% |
| | Bre5p | 8/25.8% |

Table 2.5 Mass spectrometry analysis of proteins associated at the 2nd step of splicing from Figure 2.3B.

### 2.3.7 Sedimentation comparison between the first step-arrested spliceosome and the second step arrested spliceosome

The difference in sedimentation behavior between the first step-arrested and second step-arrested spliceosomes was measured as represented by the quantitation of U6 snRNA throughout the glycerol gradient fractions (Figure 2.5). Despite the loss of more than 500 kDa with the absence of SF3, the *prp16*-arrested spliceosomes do not sediment significantly differently than spliceosomes arrested in the first step. This is likely due to the added polypeptide mass of Prp16p, Prp22p, Prp43p, and others, which in total are similar to the mass of the displaced SF3 complex and other first-step components (Figure 2.3). The similarity in mass and co-sedimentation suggest that apart from the compositional changes described above there is not a dramatic conformational change after the first step and prior to the action of Prp16p.

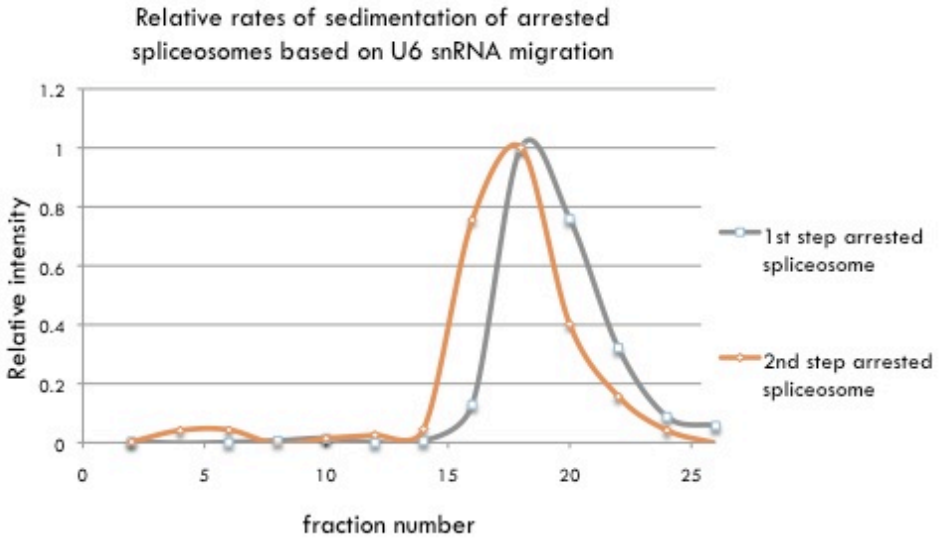


Figure 2.5 First and second step-arrested spliceosomes do not show significant differences in sedimentation.

U6 snRNA was quantitated as a representative of the second step-arrested spliceosome and measured against the sedimentation of the first step-arrested particle. The two particles do not sediment significantly differently.

### 2.3.8 Post-splicing *prp22*-arrested spliceosome characterization

After the second step, the spliced mRNA and the intron are retained within the spliceosome until the ATP-dependent function of the DEAH-box protein Prp22p releases the mRNA (Company *et al*., 1991; Schwer, 2008). To complete the DEAH-box-arrested pre-mRNA-, pre-mRNA intermediate-, or mRNA-containing spliceosome purifications spliceosomes arrested after the second step, but prior to the release of mRNA were purified using the same TAP-tagging strategy with two cold-sensitive alleles of *prp22* (DBY4172 – *prp22D613Q* and DBY4423 – *prp22G394R, D613Q*)(Noble and Guthrie, 1996). Similar to the wildtype Prp2p (Figure 2.1A) and Prp16p (Figure 2.3A) purifications, Prp22p was not associated with significant levels of snRNAs (Figure 2.6A) following affinity purification and size separation by glycerol gradient sedimentation. Spliceosomes arrested with the *prp22* cold-sensitive alleles at the Prp22p mRNA release step were affinity purified and sedimented through a glycerol velocity gradient. RNA extracted from the gradient fractions was Northern blotted for snRNAs (Figure 2.6B). The presence of U2, U5 and U6 snRNAs at ~40S of the glycerol gradient indicates the presence of a spliceosome.

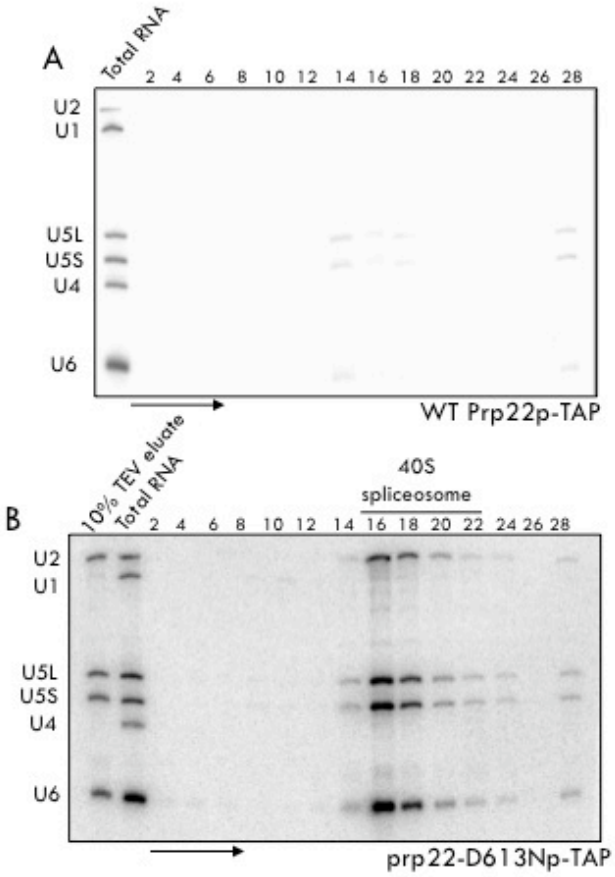


Figure 2.6 Cold-sensitive alleles of *prp22* accumulate spliceosome in the ~40S region of the glycerol gradient.

A. Prp22p-TAP affinity-purified material shows no significant accumulation of snRNAs in the glycerol gradient, after Northern blotting for all five snRNAs.

B. prp16-D613N-TAP-associated material was Northern blotted for spliceosomal snRNAs. Splicing complexes, as indicated by the presence of U2, U5 and U6, are seen in the ~40S region of the glycerol gradient. The arrow indicates the direction of sedimentation, and positions of U1, U2, U4, U5 and U6 snRNAs are shown.

### 2.3.9 Pre-mRNA characterization of a message release-arrested particle

RT-PCR analysis of the mRNA/pre-mRNA/lariat presence within the *prp22*-arrested particle was carried out as described in 2.3.5. This analysis revealed the presence of only mRNA and lariat intron, consistent with the products of the second step of splicing (Figure 2.7). No 3' splice site was detected supporting the completion of both steps of splicing, and confirming the purification of a spliceosome arrested at the release step of the splicing cycle.

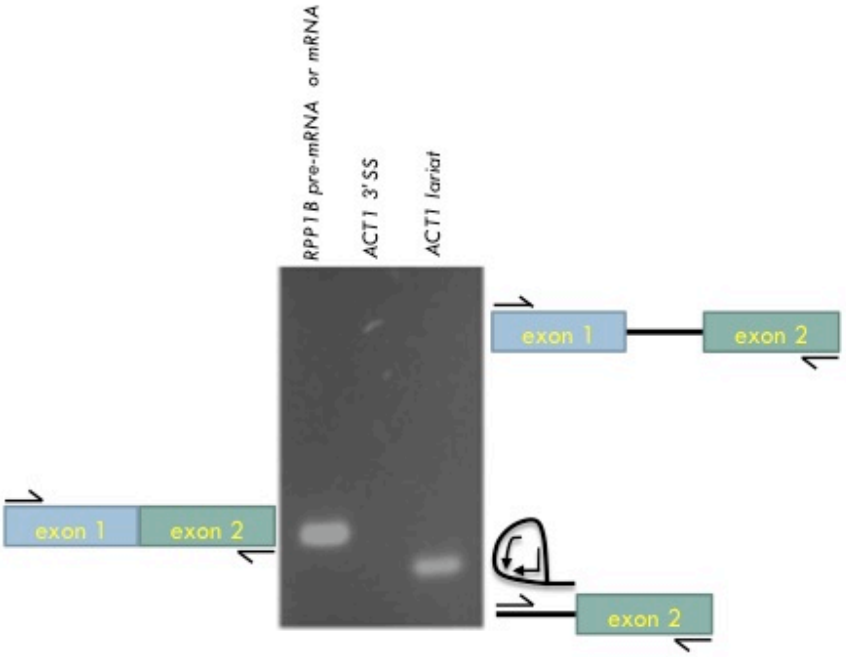


Figure 2.7 *prp22*-arrested particles contain spliced message and lariat intron as determined by RT-PCR.

RT-PCR of intron-containing transcripts was used to assess the splicing status of the intron within the *prp22*-arrested particle at NPT. The first set of primers, which anneal to the exons of *RPP1B*, show the presence of spliced message. Primers annealing to the *ACT1* intron and the 3' exon reveal the absence of a 3' splice site, which suggests both steps of splicing have come to completion. Finally, primers that are designed to amplify a lariat structure of the *ACT1* intron show that there is a lariat contained in the spliceosome. Taken together, these results confirm that this particle is arrested post-splicing.

### 2.3.10 Compositional characterization of the message release-stalled particle

MudPIT mass spectrometry analysis was once again employed to determine the composition of the arrested particle. The *prp22*-stalled particle contained a strikingly similar composition to the *prp16*-arrested spliceosome (Table 2.5). Interestingly, the second step-specific factors, Slu7p and Prp18p, were present in the message release mutant particle, consistent with their action after the Prp16p step, but prior to the second chemical step (Umen and Guthrie, 1995b). Also notable is the more complete presence of the THO/TREX complex, which is consistent with data showing that the THO/TREX complex loads late during transcription on most intron-containing transcripts, after spliceosome assembly (Abruzzi *et al*., 2004). Additionally, Yju2p and Cwc25p are absent after Prp16p action and prior to dissociation of the spliceosome. Interestingly, the presence of a deubiquitinase, Ubp3p, and its interaction partner, Bre5p, were present in both the *prp16*- and *prp22*- arrested spliceosomes. Although no published work has demonstrated a function for these proteins in pre-mRNA splicing, these factors have also been detected in post-splicing lariat RNP complexes (J Combs and SW Stevens, unpublished).

| snRNP or complex association | Protein | # of peptides identified/percentage polypeptide |
|---|---|---|
| U2 snRNP core | Lea1p | 15/59.2% |
| | Msl1p | 3/36.9% |
| U4/U6•U5 | Prp8p | 87/37.8% |
| | Brr2p | 26/16.0% |
| | Snu114p | 44/44.1% |
| U2, U5 snRNPs | Sm core | ✓ |
| NTC | Prp19p | 44/62.4% |
| | Prp46p | 33/60.5% |
| | Snt309p | 10/68.0% |
| | Cef1p | 32/48.1% |
| | Syf1p | 36/38.8% |
| | Clf1p | 27/44.7% |
| | Ntc20p | 2/35.7% |
| | Cwc2p | 19/48.1% |
| | Isy1p | 7/31.5% |
| | Syf2p | 11/40.0% |
| 2nd step factors | Prp17p | 7/20.4% |
| | Prp18p | 7/32.7% |
| | Slu7p | 13/28.3% |
| Helicase/related | Prp22p | 69/50.4% |
| | Prp43p | 22/33.3% |
| THO/TREX | Yra1p | 7/31.0% |
| | Tex1p | 5/16.4% |
| | Hpr1p | 6/13.2% |
| | Sub2p | 3/8.5% |
| | Thp2p | 2/13.4% |
| | Mft1p | 3/10.5% |
| mRNA binding | Sto1p | 16/22.6% |
| | Cbc2p | 4/31.2% |
| Spliceosome associated | Bud31p | 5/22.3% |
| | Prp45p | 21/55.7% |
| | Cwc22p | 21/42.6% |
| | Cwc15p | 9/45.1% |
| | Cwc21p | 4/20.0% |
| | Ecm2p | 15/37.6% |
| Novel | Ubp3p | 16/30.6% |
| | Bre5p | 13/36.7% |

Table 2.6 Mass spectrometry analysis of the associated proteins from message release-arrested spliceosomes from Figure 2.6B.

### 2.3.11 Sedimentation comparison between the first step-arrested spliceosome, the second step-arrested spliceosome, and the message release-arrested spliceosome

Quantitation of U6 snRNA from each of the fraction of the arrested spliceosome purifications (Figure 2.1B, 2.3B, and 2.6B) was performed and plotted across the glycerol gradients. The post-splicing, *prp22*-arrested spliceosomes reproducibly sedimented three fractions higher in the gradient than either the *prp2*- and *prp16*-arrested spliceosomes (Figure 2.8). Because the protein composition of this particle is similar to that of the *prp16*-arrested particle, this repeatable, significant shift in spliceosome sedimentation indicates a conformation change following the second step of chemistry, resulting in a slower migrating particle.

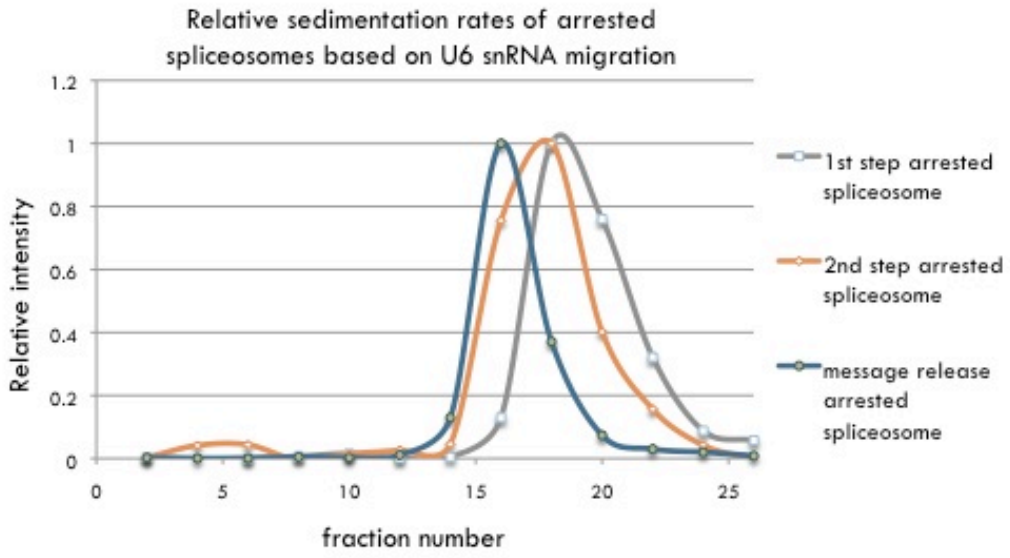


Figure 2.8 The sedimentation pattern of *prp2*-, *prp16*- and *prp22*-arrested spliceosomes was compared and suggests a measureable conformation change after the second step of splicing.

U6 snRNA was quantitated as a representative of the arrested spliceosomes and measured for sedimentation through a glycerol gradient. Although *prp2*- and *prp16*-arrested particles do not sediment significantly different, *prp16*-arrested spliceosomes sediment approximately 3 fractions removed compared to the *prp22*-arrested particle, suggesting a substantial rearrangement after the two steps of splicing are completed.

## 2.4 DISCUSSION

Many important discoveries in the field of pre-mRNA splicing have resulted from compositional comparison of (1) constituent parts of the spliceosome (Caspary and Seraphin, 1998; Caspary *et al*., 1999; Neubauer *et al*., 1997; Stevens and Abelson, 1999; Stevens *et al*., 2001; Stevens *et al*., 2002), (2) assembled or partially assembled functional complexes (Bessonov *et al*., 2008; Chen *et al*., 2007; Hartmuth *et al*., 2002; Jurica and Moore, 2002; Zhou *et al*., 2002), and (3) products of the reaction (Makarov *et al*., 2002). In this dissertation work, spliceosomes arrested *in vivo* at three helicase checkpoints during the catalytic phase of the splicing cycle have been purified using gentle, but stringent affinity chromatography techniques. Due to the differences likely to be encountered throughout the assembly of the spliceosome *in vitro* upon a single transcript versus the co-transcriptional assembly of the spliceosome *in vivo*, we believe that the data presented here accurately represent the composition and functionality of these complexes.

Our mass spectrometry data from each of the first step-, second step- and message release-arrested spliceosomes (as shown side-by-side in Table 2.7) show that one striking feature among each of the arrested spliceosomes is the absence of tri-snRNP-associated factors, with the exceptions of the Sm core, Prp8p, Brr2p and Snu114p, which are present throughout the assembly and splicing. Our data show that the majority of the tri-snRNP associated factors are dispensable during the catalytic steps of splicing. This result differs slightly from that of the activated human spliceosomes (Makarov *et al*., 2002), in which other tri-snRNP components were detected. This discrepancy may be due to the operational differences in purifying synchronized splicing complexes formed *in vivo*

versus those formed *in vitro*, or less likely, might reflect functional differences in the two systems.

More striking is that the SF3a and SF3b components are completely absent after the first step. Other differences between *prp2*- and *prp16*-arrested spliceosomes include the addition of factors known to participate in the second step of splicing (e.g. Prp16p, Prp22p), and others that have been implicated in pre-mRNA splicing by association with spliceosomes (e.g. several Cwc proteins), but for which there are not yet data demonstrating a function in splicing. Interestingly, THO/TREX complex components appear only after the first step of splicing and persist through the second step. This temporal difference is consistent with data showing that the THO/TREX complex loads onto intron-containing transcripts subsequent to assembly events (Abruzzi *et al*., 2004).

We also note that the entire RES complex is absent after the first step of splicing. The heterotrimer has been implicated in the retention of unspliced pre-mRNA in the nucleus and indeed, tagged RES components only co-purify radiolabeled pre-mRNA and no other pre-mRNA splicing intermediates (Dziembowski *et al*., 2004), consistent with our mass spectrometry results. How the RES complex might enforce a nuclear retention phenotype when it is only associated prior to the first step is not currently understood. Other differences between the *prp2*- and *prp16*-arrested spliceosomes include the addition of the post-splicing Prp43p helicase (Arenas and Abelson, 1997). While Prp43p is not known to be required for the second step of for mRNA release, its presence indicates that it may either have an unknown role at these steps, or that its presence is required or tolerated without playing a catalytic role.

Compositional differences between the *prp16*- and *prp22*-arrested spliceosomes are relatively minimal; however, they are consistent with previous reports. These include the addition of Prp18p and Slu7p in the *prp22*-arrested spliceosome, which is consistent

with the proposed role for these second step factors. Both have been shown to bind and function after the ATPase-dependent function of Prp16p (Jones *et al*., 1995; Umen and Guthrie, 1995b). In addition, Prp17p, is present throughout the splicing cycle according to our data and data from the Vijayraghavan laboratory (Sapra *et al*., 2008). Several THO/TREX complex components are stably bound after the second step; however, none of the isolated spliceosome contains the largest THO/TREX component, Tho2p.

Other differences include the addition of Yju2p and Cwc25p after the Prp2p step, and their removal after the second step. Previous work from the Cheng and colleagues show these factors function prior to the first step (Chiu *et al*., 2009; Liu *et al*., 2007). Other recent data from the Lührmann laboratory show the stimulatory effect of Cwc25p during the first step of splicing (Warkocki *et al*., 2009). Our data demonstrates that *in vivo*-assembled first step-arrested spliceosomes do not contain these proteins suggesting that these factors are added after the Prp2p step or that they are only required for splicing a subset of transcripts.

| | | *prp2* 1st step block | *prp16* 2nd step block | *prp22* post-splicing |
|---|---|---|---|---|
| U2 snRNP core | Lea1p | ✓ | ✓ | ✓ |
| | Msl1p | ✓ | ✓ | ✓ |
| SF3b | Rse1p | ✓ | | |
| | Hsh155p | ✓ | | |
| | Cus1p | ✓ | | |
| | Hsh49p | ✓ | | |
| | Ysf3p | ✓ | | |
| | Rds3p | ✓ | | |
| SF3a | Prp21p | ✓ | | |
| | Prp11p | ✓ | | |
| | Prp9p | ✓ | | |
| U4/U5•U6 snRNP | Prp8p | ✓ | ✓ | ✓ |
| | Brr2p | ✓ | ✓ | ✓ |
| | Snu114p | ✓ | ✓ | ✓ |
| U2, U5 snRNPs | Sm core | ✓ | ✓ | ✓ |
| NTC | Prp19p | ✓ | ✓ | ✓ |
| | Prp46p | ✓ | ✓ | ✓ |
| | Snt309p | ✓ | ✓ | ✓ |
| | Cef1p | ✓ | ✓ | ✓ |
| | Syf1p | ✓ | ✓ | ✓ |
| | Clf1p | ✓ | ✓ | ✓ |
| | Ntc20p | ✓ | ✓ | ✓ |
| | Cwc2p | ✓ | ✓ | ✓ |
| | Isy1p | ✓ | ✓ | ✓ |
| | Syf2p | ✓ | ✓ | ✓ |
| 2nd step factor | Prp17p | ✓ | ✓ | ✓ |
| | Prp18p | | | ✓ |
| | Slu7 | | | ✓ |

| | | *prp2* 1st step block | *prp16* 2nd step block | *prp22* post-splicing |
|---|---|---|---|---|
| Helicase/related | Prp2p | ✓ | | |
| | Spp2p | ✓ | | |
| | Prp16p | | ✓ | |
| | Prp22p | | ✓ | ✓ |
| | Prp43p | | ✓ | ✓ |
| RES complex | Bud13p | ✓ | | |
| | Pml1p | ✓ | | |
| | Ist3p | ✓ | | |
| mRNA binding | Sto1p | ✓ | ✓ | ✓ |
| | Cbc2p | ✓ | ✓ | ✓ |
| THO/TREX | Yra1p | | ✓ | ✓ |
| | Tex1p | | | ✓ |
| | Hpr1p | | ✓ | ✓ |
| | Sub2p | | | ✓ |
| | Thp2p | | ✓ | ✓ |
| | Mft1p | | ✓ | ✓ |
| Spliceosome associated | Yju2p | | ✓ | |
| | Bud31p | ✓ | | ✓ |
| | Prp45p | ✓ | ✓ | ✓ |
| | Cwc22p | ✓ | ✓ | ✓ |
| | Cwc24p | ✓ | | |
| | Cwc27p | ✓ | | |
| | Cwc25p | | ✓ | |
| | Cwc15p | | ✓ | ✓ |
| | Cwc21p | | ✓ | ✓ |
| | Ecm2p | ✓ | ✓ | ✓ |
| Novel | Ubp3p | | ✓ | ✓ |
| | Bre5p | | ✓ | ✓ |

Table 2.7 Compositional comparison between arrested spliceosomes.

MudPIT mass spectrometry analysis of first step (*prp2*)-arrested, second step (*prp16*)-arrested and message release (*prp22*)-arrested spliceosomes. The most striking difference between the three purified catalytically blocked spliceosomes is the lack of SF3a and SF3b complexes following the Prp2p step. Other notable features include the lack of tri-snRNP proteins other than Brr2p, Prp8p and Snu114p, and the specific presence of Yju2p and Cwc25p only before the second chemical step.

These data represent the first purification of catalytically synchronized, arrested particles. Using the cold-sensitive nature of the three DEAH box factors required for the catalytic steps of splicing and the release of the spliced message, we have accumulated spliceosomes, which exist only in a transient nature under wildtype conditions. Upon shift of the cold-sensitive strains to non-permissive temperatures, the mutant DEAH-box proteins are prevented from carrying out their required functions, thus arresting growth of cells, and presenting a means by which the particles can be biochemically characterized. In addition, we have shown that the defect in the first step-arrested particle is reversible and rescued by raising the temperature to where prp2G551Np is functional and adding ATP. This presents the opportunity for functional characterization of the first step of splicing through "chasing" the particle through the first step of splicing. The following chapter will elaborate on the compositional differences observed between the first step- and second step-arrested spliceosomes by functional testing.

# Chapter 3: Removal of the SF3 complex initiates the first step of splicing

## 3.1 Background: Authorization for the first step of splicing

### 3.1.1 Spliceosome assembly around the branchpoint sequence

One of the earliest events in spliceosome assembly includes the association of the U2 snRNA with the branchpoint sequence of the pre-mRNA (Parker *et al*., 1987) (see Illustration 3.1). The conserved uridine residue of the U2 snRNA that is situated across from a dinucleotide of adenosine residues of the branchpoint sequence (see Illustration 3.1) is modified to a pseudouridine in all eukaryotes (Patton *et al*., 1994; Yu *et al*., 1998). This modification is thought to position the 2'-OH of the branchpoint adenosine accessibly for nucleophilic attack (Newby and Greenbaum, 2002). This is consistent with the idea that the branchpoint nucleophile is activated by being bulged from the U2 snRNA/pre-mRNA duplex (Query *et al*., 1994). Furthermore, in the absence of a conserved 5' splice site, the branchpoint nucleophile has been shown to attack the branchpoint binding sequence of the U2 snRNA, demonstrating the highly reactive nature of the nucleophile (Smith *et al*., 2007) most likely due to its position within the active site of the spliceosome. The accessibility and presumably early "activation" of the branchpoint adenosine suggest that a mechanism must serve to prevent premature attack of the 5' splice site in the events leading up to the first chemical step of splicing.


Illustration 3.1 U2 snRNA base pairs with the pre-mRNA early in spliceosome assembly.

The conserved uridine residue, underlined in the illustration, is modified to pseudouridine, which is thought to assist in positioning, or bulging, the branchpoint adenosine residue of the pre-mRNA into a conformation where it can elicit its nucleophilicity.

Formation of the duplex between the U2 snRNA and the branchpoint sequence occurs concomitantly with the joining of the U2 snRNP-associated proteins, contained within two subcomplexes known as SF3a (Brosi *et al*., 1993a; Brosi *et al*., 1993b) and SF3b (Will *et al*., 2002). These complexes interact upstream of and on the branchpoint sequence, respectively, and are thought to anchor the U2 snRNP to the pre-mRNA (Gozani *et al*., 1996; Ruby *et al*., 1993). These factors are classically known for their identification of and interaction with the branchpoint, as the U2 snRNP contacts the pre-mRNA in spliceosome assembly; however, they remain part of the spliceosome throughout activation. Their purpose in events following spliceosome assembly had not been hypothesized upon or investigated prior to this work.

### 3.1.2 Uncoupling Prp2p function from the first step of splicing

As described in Chapter 2, Prp2p produces an ATP-dependent rearrangement that is required for the first step of catalysis; however, this remodeling event has been demonstrated to be followed up by the actions of at least two other factors, which are proposed to also act prior to the first chemical step of splicing. Recent *in vitro* reconstitution studies from the Cheng laboratory have suggested two novel factors in the first step of splicing. They propose that Yju2p acts in an ATP-independent, Prp2p-independent manner to promote the first step of splicing after Prp2p function (Liu *et al*., 2007), while Cwc25p relies on the functional rearrangement by Prp2p and the presence of Yju2p to elicit function (Chiu *et al*., 2009). The Lührmann laboratory has also recently demonstrated the functioning of Cwc25p in stimulating the first step of splicing(Warkocki *et al*., 2009), however, their results do not conclusively identify

Cwc25p as essential in the first step. Though *in vitro* studies have placed the novel factors as acting in the first step of splicing, their mechanisms of action have not been determined.

Our *in vivo*-arrested first step particles (as described in Chapter 2) provided a system by which we can functionally analyze the first step of splicing. We found that Prp2p functions in an ATP-dependent manner to remove the SF3 complex from the U2/pre-mRNA duplex, only after an ATP-independent step induces a conformational rearrangement. We propose that this ATP-independent conformational restructuring results in a spliceosomal configuration, which is optimal for the chemistry of the first step of splicing. Prp2p then induces an ATP-dependent release of the complex that has been previously shown to intimately interact with the branchpoint sequence, likely serving to sequester the highly reactive nucleophile from premature attack. The evidence from our functional analysis of *in vivo* purified spliceosomes construct a model, which includes these mechanistic events in the initiation of the first step of splicing.

## 3.2 MATERIALS AND METHODS

### 3.2.1 Tagging vectors

| Vector | Purpose | Marker |
|---|---|---|
| pFA6a-13Myc-His3MX6 | myc tagging | *HIS3* |
| pFA6a-kanMX6-PGAL-3HA | N-terminal HA tag, galactose inducible promoter | *KANr* |

Table 3.1 Vectors used for creating C-terminal tags in Chapter 3.

### 3.2.2 Yeast strains

| Strain | Mating type | Genotype |
|---|---|---|
| SS1353 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*PRP2-TAP::URA3, PRP9MYC::HIS3* |
| SS1354 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*prp2G551N-TAP::URA3, HSH155MYC::HIS3* |
| SS1357 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*PRP2-TAP::URA3, PRP9MYC::HIS3, KAN::GAL-HA-YJU2* |
| SS1358 | a | *ura3, leu2, trp1, his3, lys2, prp2::TRP,* pRS415-*PRP2-TAP::URA3, HSH155MYC::HIS3, KAN::GAL-HA-YJU2* |

Table 3.2 Yeast strains used in Chapter 3.

### 3.2.3 Oligos

| Oligo name | Sequence |
|---|---|
| PRP9F2 | 5'GAAGAAGGAAATGTAATGAGTAAGAAGGTCTACGATGAACTTAAGAA GCAAGGTTTGGTGCGGATCCCCGGGTTAATTAA3' |
| PRP9R1 | 5'AAGGTCATTACTAAGAATAGTATAAATACATACAACTGCTATCTATCA AACAAATATACAGAATTCGAGCTCGTTTAAAC 3' |
| HSH155F2 | 5'TTTTACCCCGTTACACCAGACAACAATGAAGAATATATAGAAGAACTG GATTTAGTTCTGCGGATCCCCGGGTTAATTAA3' |
| HSH155R1 | 5'GCATCGTGACGTGATGATATAGGTGTGTCAAGTAAAATATTCTTACAA GTTGTGGTTATTGAATTCGAGCTCGTTTAAAC 3' |
| PRP2TAPA | 5'CACAAATCTTTAAAGATTTAATTGACGATAAAACAAATAGGGGGAGGC GGTCCATGGAAAAGAGAAG3' |
| PRP2TAPB | 5'TGCATATAGAATGGAGCCTGCGTTTCTAGCAATACACATACACCTGTC AATACGACTCACTATAGGG3' |
| YJU2F4 | 5'AAAAAGCAAGAAATCACTACGGTGACCTCAGTAGAACGAACTAATAA GTCAAACCAAATTGAATTCGAGCTCGTTTAAAC 3' |
| YJU2R3 | 5'AGCTTCCAGTGGGTTGTAATCCGGTGGATAGTACTTGTTAATAGCTTTT CTTTCAGAGCACTGAGCAGCGTAATCTG3' |
| U1 probe | 5'GAATGGAAACGTCAGCAAACAC3' |
| U2 probe | 5'AAGAACAGATACTACACTTGA3' |
| U4 probe | 5'ACCATGAGGAGACGGTCTGG3' |
| U5 probe | 5'ATGTTCGTTATAAGTTCTATAGGC'3 |
| U6 probe | 5'AGGGGAACTGCTGATC3' |
| RPP1B ex1SS5 | 5'ATTGACTATCACCAAGGCCGC3' |
| RPP1B ex2-3 | 5'GTGACGCTGCTGCTGAAGAAGAAA3' |
| YFL039C_1_p_F | 5'AGGGGCTTGAAATTTGGAAAAA 3' |
| YFL039C_1_L_R | 5'GCAAGCGCTAGAACATACATAGTACA3' |
| YFL039C_1_I_F | 5'TTGCTTCATTCTTTTTGTTGCT3' |
| YFL039C_1_I_R | 5'GCAAAACCGGCTTTACACAT3' |

Table 3.3 Oligos used for PCR of tags, Northern blotting, and RT-PCR in Chapter 3.

### 3.2.4 Strain creation

Strains listed in Table 3.2 were created using the PCR generated fragment strategy described in Chapter 2, section 2.2.4. For C-terminal myc tags, the tag was amplified from pFA6a-13Myc-His3MX6, which contains a *HIS* selectable marker. For N-terminal, Galactose drive, HA-tags, fragments were amplified from pFA6a-kanMX6-PGAL-3HA, which has a *KANr* selectable marker 5' to the tag.

Transformants were screened, as described in Chapter 2, section 2.2.4, except using antibodies to recognize the myc tag. Whole cell extract was then separated by polyacrylamide gel electrophoresis, and transferred to nitrocellulose (Bio-Rad) membrane overnight at 25V at 4°C in transfer buffer (20% methanol, 0.001% SDS, 386 mM glycine, 48 mM Tris base). The membrane was then subjected to Western blotting for the myc tag as follows: the membrane was blocked for 1 hour at room temperature in BLOTTO, washed twice for 5 minutes each in PBST, incubated for 1 hour in BLOTTO plus a 1:5000 dilution of primary antibody (mouse α-myc) or a 1:1000 dilution of anti-HA (Covance), washed 4x for 5 minutes in PBST, then incubated in BLOTTO plus a 1:5000 dilution of goat-α-mouse HRP-conjugated secondary (Rockland) for both the HA and the myc Westerns for 1 hour, and finally washed four times for fifteen minutes each in PBST. Chemiluminescent detection was performed to detect the western blot signal (Perkin Elmer).

### 3.2.5 Metabolic depletion of *YJU2*

Strains containing a metabolically depletable *YJU2* gene were grown in YP + 2% galactose at permissive temperature (PT) to OD ($A_{600}$) = 0.5, washed in prewarmed YPD, resuspended in fresh YPD, and allowed to grow for 3 hours at PT. Cell were then shifted to 25°C for 45 minutes before harvesting. Cells were harvested as described in Chapter 2, section 2.2.6.

### 3.2.6 Functional analysis of arrested spliceosomes

Arrested spliceosomes were purified, as described in section 2.2.7. Following eluting with TEV protease, purified material was divided into 4 equal aliquots and treated for 20 minutes under the following conditions (Stevens and Abelson, 2002):

1. 60 mM $KH_2PO_4$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine at PT in the presence of ATP
2. 60 mM $KH_2PO_{4,}$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine at PT in the absence of ATP
3. 60 mM $KH_2PO_{4,}$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine on ice in the absence of ATP
4. 60 mM $KH_2PO_{4,}$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine on ice in the presence of the non-hydrolyzable substrate, AMP-PNP

Splicing reactions were layered on 10-30% glycerol gradients and centrifuged as described in Chapter 2, section 2.2.7.

### 3.2.7 Western blotting for MYC tag across the glycerol gradients

Proteins from even fractions of the glycerol gradient were phenol/chloroform extracted and precipitated in the presence of 10 μg of BSA (NEB). Protein was separated by SDS-PAGE, transferred, and Western blotting for the myc tag was conducted as described in section 3.2.4.

### 3.2.8 Northern blotting for snRNAs

RNA extracted from the glycerol gradients was precipitated and analyzed by Northern blotting for all five spliceosomal snRNAs as described in Chapter 2, section 2.2.8.

### 3.2.9 RNA quantitation and normalization

Quantitation of U6 snRNA in each Northern blot was performed as described in Chapter 2, section 2.2.9.

### 3.2.10 Native gel electrophoresis of splicing reactions

Native gel electrophoresis of splicing reactions using affinity-purified spliceosomes was performed using 20 cm X 20 cm X 0.75 mm gels (0.5X TAE and 4% 80:1 polyacrylamide). RNA was separated by gel electrophoresis at 250 V for 11.75 hours in 0.5X TAE running buffer at 4°C, with recirculating buffer. Gels were transferred to nylon membrane and analyzed by Northern blotting for U6 snRNA as described in Chapter 2, section 2.2.8.

### 3.2.11 Spliceosome pelleting experiment

TEV elutions from purifications of arrested first-step spliceosomes were layered on 10-30% glycerol gradients under conditions described in Chapter 2, section 2.2.7. The peak spliceosome fraction from the glycerol gradient was split into three equal samples and treated for 20 minutes as follows:

1. 1 M NaCl
2. 60 mM $KH_2PO_4$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine on ice in the absence of ATP
3. 60 mM $KH_2PO_4$, 3.2 mM $MgCl_2$, 3% PEG-5000, 1 mM spermidine at PT in the presence of ATP

Samples were layered onto a 30% glycerol cushion containing 20 mM HEPES, pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 1 mM DTT, 0.2 mM PMSF, and centrifuged at 80,000 RPM for 80 minutes in a TLA 100.3 rotor at 4°C. The supernatant was removed and the pellet was resuspended. Extracted protein was analyzed by Western blotting for the myc tag (as described in Chapter 3, section 3.2.4)

### 3.2.12 Protein quantitation

Quantitation of proteins in Western blots was performed using a NightOWL II LB 983 (Berthold Technologies) by ultracold-CCD camera detection. Signal intensities were integrated over the lanes of the Western blot and values for each glycerol gradient fraction was normalized against the lane with the most intense signal, which was give a value of 1. Automatic average background subtraction was performed for each Western blot independently.

### 3.2.13 RT-PCR

Extracted and precipitated RNA from each of the splicing reactions (described in Chapter3, section 3.2.6, samples 1-3) was subjected to RT-PCR (as described in Chapter 2, section 2.2.10) to detect pre-mRNA/mRNA (*RPP1B*), 3' splice site (*ACT1*), or lariat structure (*ACT1*).

## 3.3 RESULTS

### 3.3.1 *In vivo* depletion of Yju2p

Although Prp2p function initiates the first step of splicing, the recently described factor Yju2p was reported to be required after Prp2p function but prior to chemistry (Liu *et al*., 2007). Yju2p is not detectable by mass spectrometry in *prp2*-arrested particles (Table 2.4); however to ensure that we could confidently distinguish between the function of Prp2p and Yju2p in the arrested spliceosomes, we metabolically depleted Yju2p from prp2G55Np-TAP cells by repressing the *GAL*-driven, HA-tagged Yju2p. As shown in the Figure 3.1, Yju2p is not detectable by western blot after a 90-minute shift to dextrose.

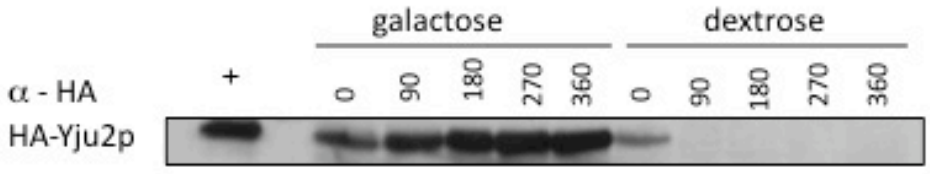


Figure 3.1 Yju2p can be metabolically depleted from prp2G551Np-TAP cells.

Western blot analysis of GAL-HA-Yju2p through a time course in galactose (*left*) or after shift to dextrose (*right*) shows that Yju2p is not detectable after 90 minutes in media containing the repressive carbon source, dextrose.

### 3.3.2 Functional analysis of *prp2*-arrested spliceosomes by Northern blotting

First step-arrested spliceosomes, from cells depleted of Yju2p and containing a MYC tag on Prp9p, were isolated by affinity chromatography using the TAP tag on prp2G551Np. Prp9p is the yeast homolog of the mammalian SF3a60 factor, which is a component of the U2 snRNP-associated trimeric SF3a complex. Purified, arrested spliceosomes were incubated under splicing conditions at NPT with no ATP treatment, under splicing conditions at PT with no ATP treatment, under splicing conditions at PT with AMP-PNP treatment, or under splicing conditions at PT with ATP. After incubation, the reactions were centrifuged through spliceosome-resolving glycerol velocity gradients, as described in Chapter 2, section 2.2.6. Following fractionation, the RNA from the even fractions from the glycerol gradients was subjected to Northern blot analysis to detect spliceosomal snRNAs. Quantitation of U6 snRNA from each of the particle treatments is shown in Figure 3.2.

Spliceosomes incubated at NPT (Figure 3.2) sedimented identically to those from an untreated sample in which Yju2p was not depleted (Figure 2.1B). In samples treated at PT without ATP, the peak of the U6 snRNA signal was shifted three fractions higher in the gradient (Figure 3.2), indicating a substantial change in shape or in composition. This suggests that a novel ATP-independent event may be catalyzed by Prp2p itself and represents a conformational change. This evidence is corroborated by the treatment of the particle at PT with the non-hydrolyzable substrate, AMP-PNP, which shows the same shift in sedimentation as seen in the absence of ATP. Furthermore, incubation at PT in the presence of ATP shows no additional shift in glycerol gradient sedimentation of the spliceosome, which suggests there are no further significantly detectable changes in conformation upon ATP hydrolysis.

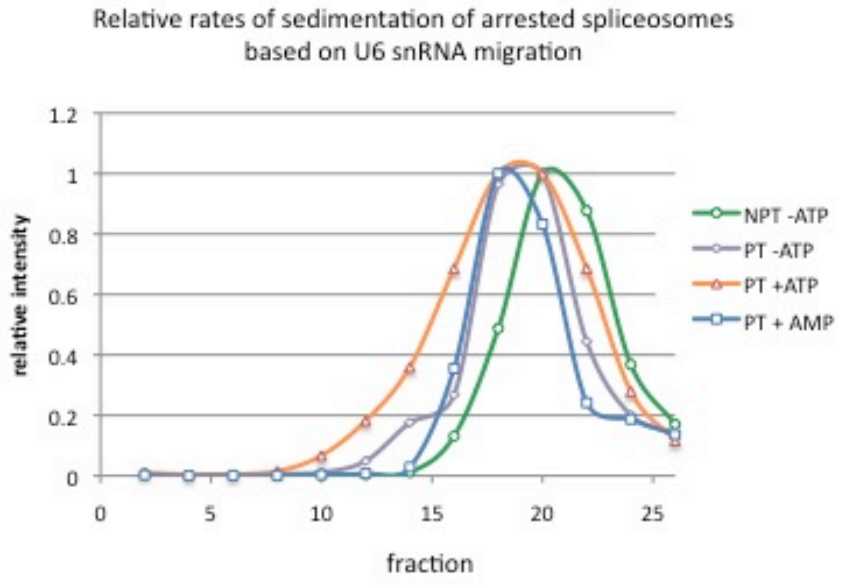


Figure 3.2 Spliceosome sedimentation, as quantitated by U6 snRNA, suggests an ATP-independent role for Prp2p.

RNA from glycerol gradients was subject to Northern blot analysis for spliceosomal RNAs. U6 was quantitated and normalized. Upon treatment at PT for the *prp2* mutant allele, spliceosomal sedimentation is shifted three fractions higher in the glycerol gradient, suggesting an ATP-independent role for Prp2p.

### 3.3.3 Functional analysis of *prp2*-arrested spliceosomes by Western blotting for the Prp9-myc

Protein was precipitated following treatment as described in 3.2.6 and glycerol gradient sedimentation as described in Chapter 2, section 2.2.7, and subjected to Western blot analysis to detect Prp9p-myc. Treatment at NPT in the absence of ATP demonstrates that Prp9p cosediments with the spliceosome (Figure 3.3A and E and Figure 3.2). Similarly, upon treatment at PT in the absence of ATP (Figure 3.3B) or the presence of the AMP-PNP at PT (Figure 3.3C and E and Figure 3.2), Prp9p remains in spliceosome-sized fractions. However, when incubated at PT in the presence of ATP, Prp9p-myc is detected near the top of the gradient (Figure 3.3D and E and 3.2), suggesting Prp9p is displaced from the spliceosome upon ATP hydrolysis by Prp2p. A quantitative analysis of the sedimentation behavior of SF3 (Figure 3.3E) and spliceosomes from Figure 3.2, reveals that the displacement of SF3 from the spliceosome does not affect the sedimentation of the spliceosome.

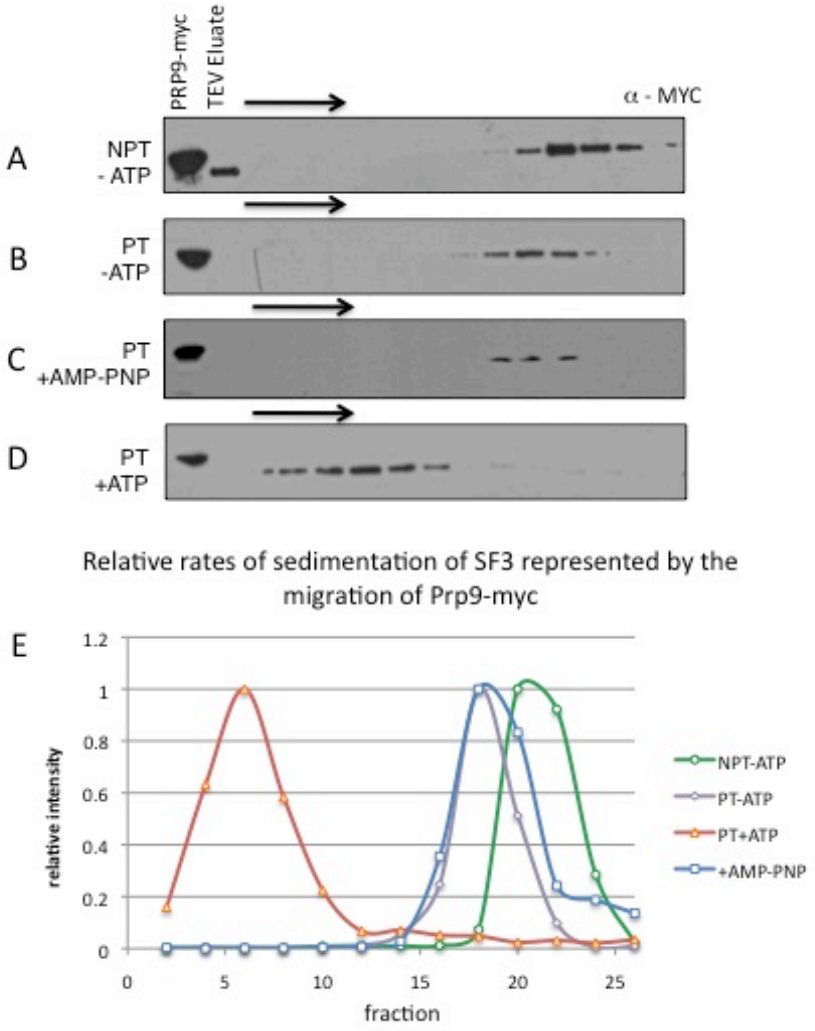


Figure 3.3 Upon ATP treatment, the SF3 complex is displaced from the spliceosome.

First step-arrested spliceosomes were treated at NPT in the absence of ATP (A), PT in the absence of ATP (B), PT in the presence of AMP-PNP (C), or PT in the presence of ATP (D). As shown in D and E, Prp9p is displaced from the spliceosome upon treatment with ATP implicating a role for ATP-hydrolysis by Prp2p in removal of SF3.

### 3.3.4 Native gel electrophoresis of treated first step spliceosomes

To further demonstrate that the shifts in glycerol gradient sedimentation seen for first step-arrested spliceosomes treated at PT correspond to authentic changes in shape, native gel electrophoresis of the treated particles was performed (Figure 3.4). Upon treatment at PT with no ATP (lane 2), treatment at PT with ATP (lane 3), and treatment at PT with AMP-PNP (lane 4) the migration of spliceosomes is faster than the spliceosomes treated at NPT (lane 1). This migration difference corresponds to the sedimentation shift seen in Figure 3.2, where spliceosomes treated at PT run three fractions higher in the glycerol gradient than those incubated at NPT. Thus, the migration difference is dependent upon raising the arrested *prp2* particle to permissive temperatures, suggesting a conformational change is induced by Prp2p. The results from both Figures 3.2 and 3.4 are consistent with an ATP-independent conformational change occurring concurrently with relieving the cold-sensitive *prp2* defect, and suggest a conformational rearrangement is induced by Prp2p.

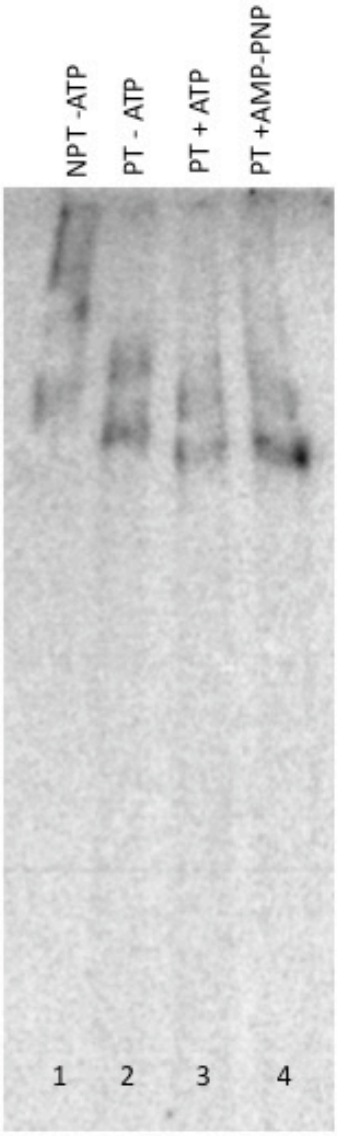


Figure 3.4 Native gel electrophoresis of treated particles from affinity purified *prp2*-arrested spliceosomes confirms the size shift, suggesting a conformational change upon treatment of the particle at PT.

Affinity elutions were split into four samples and treated as indicated. Comparison of the migrations between the untreated spliceosome (lane 1) and the samples treated at PT (lanes 2, 3, and 4) show the shift in spliceosome size, suggesting an ATP-independent conformational change.

### 3.3.5 Characterization of pre-mRNA within functional 1st step spliceosomes

To demonstrate the functionality of the arrested spliceosomes, RT-PCR was performed on RNAs contained in representative peak fractions from the glycerol gradients of the affinity-purified, treated spliceosomes (Figure 3.2). Using the strategy described in Chapter 2, sections 2.3.5 and 2.3.9, to detect pre-mRNA/mRNA, the 3' splice site, and the lariat structure, RT-PCR analysis conclusively demonstrates the functionality of the first-step arrested spliceosomes.

Pre-mRNA is exclusively found in the untreated, arrested spliceosomes, as seen in lane 4 in Figure 3.5. This is confirmed by the presence of the 3' splice site, as shown in lane 5, and the absence of the lariat, shown in lane 6. Upon treatment of the arrested spliceosome at PT in the absence of ATP pre-mRNA is exclusively found (lane 1), corroborated by the presence of 3' splice site (lane 2) and absence of lariat (lane 3). Finally, RT-PCR analysis of the spliceosome at PT in the presence of ATP confirms that the presence of ATP drives the first chemical step, as seen by the absence of pre-mRNA or mRNA (lane 7), the presence of the 3' splice site (lane 8) and the presence of the lariat structure (lane 9). The band corresponding to the 3' splice site (lane 8) and the lariat structure (lane 9) in addition to the absence of pre-mRNA (lane 7) indicates the products of the first step of chemistry. These data confirm the functionality of the *prp2*-arrested particle, and demonstrate that the first step of splicing is dependent upon the presence of ATP, which as described in the previous sections is concurrent with SF3 displacement.

The lack of first step products seen following treatment of the particle at permissive temperature also indicates that the ATP-independent conformational change is not coincident with or sufficient to promote the first chemical step. In the presence of ATP at PT and in the absence of Yju2p, the vast majority of pre-mRNA was chased to

lariat intermediate, demonstrating that even if Yju2p is required for the first step of splicing for a subset of introns, it is not required for the displacement of SF3 under these conditions, as SF3 is not spliceosome associated after Prp2p function (Table 2.5 and Figure 3.3).

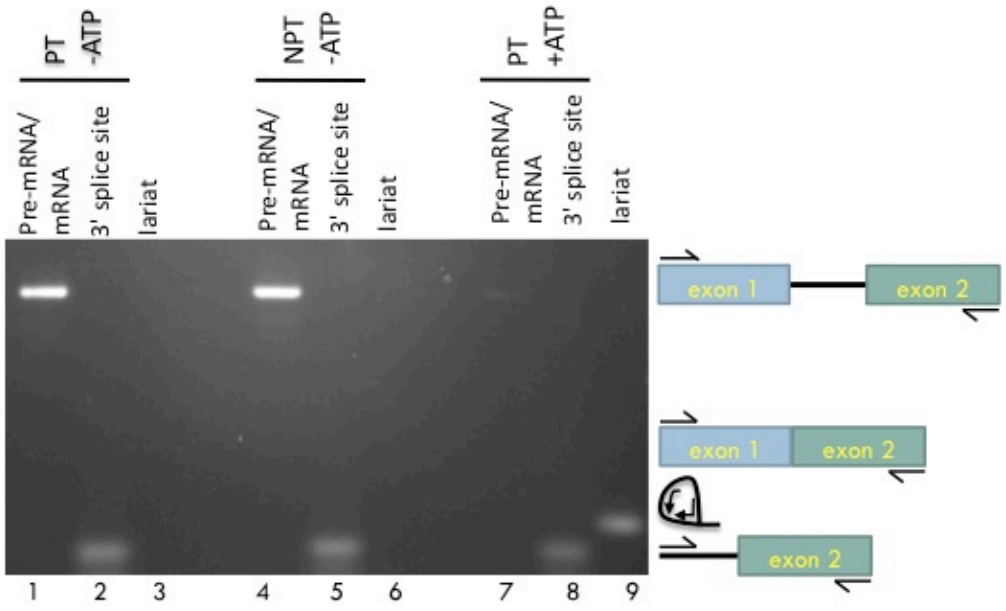


Figure 3.5 *prp2*-arrested particles are functional, and upon treatment with ATP, can be chased through the first step of splicing.

When affinity purified *prp2*-arrested spliceosomes were left untreated, presence of pre-mRNA (lane 4), presence of 3' splice site (lane 5), and absence of lariat structure (lane 6) indicate a first step chemical block. Upon elevating the temperature of the spliceosome to PT, presence of pre-mRNA (lane 1), presence of 3' splice site (lane 2) and absence of lariat structure once again confirm a first step block. Treatment with ATP at PT chases the pre-mRNA to the products of the first chemical reaction, as shown by the absence of pre-mRNA (lane 7), presence of 3' splice site (lane 8) and presence of the lariat structure (lane 9).

### 3.3.6 SF3b displacement is independent of Yju2p

SF3a and SF3b are both present in the assembled, *prp2*-arrested spliceosomes (Table 2.4); however, since each subcomplex can be isolated independently they may behave very differently in terms of their release from the spliceosome. To test for the displacement of both SF3a and SF3b, a glycerol-cushion pelleting assay to test for SF3 release was employed. Conditions designed to pellet intact spliceosomes, but incapable of pelleting displaced SF3a or SF3b were used. First step-arrested spliceosomes containing either Prp9p-myc (SF3a component) or Hsh155p-myc (SF3b component) were affinity purified. To enrich for homogeneous populations, spliceosomes were further purified by glycerol velocity gradient sedimentation, as described in Chapter 2, section 2.2.7. The peak fraction of spliceosome particles from the glycerol gradient was evenly divided into three equal samples and treated for 20 minutes either under splicing conditions at NPT in the absence of ATP, under splicing conditions at PT in the presence of ATP, or with 1 M NaCl. The reactions were sedimented through a glycerol cushion to separate the spliceosome from released factors. Proteins extracted from the supernatant and the pellet were analyzed for the presence of the myc tag by Western blotting. The SF3a and SF3b complexes are known to be sensitive to elevated salt levels (Behrens *et al*., 1993; Bessonov *et al*., 2008); thus, the salt-treated samples serve as a control for displacement of SF3 from the purified spliceosome (Figures 3.6A and B, lanes 7 and 8). At NPT, they are found exclusively in the pellet (Figures 3.6A and B, lanes 3 and 4). Both SF3a and SF3b are displaced from the conditions, as shown by the presence of Prp9p-myc and Hsh155p-myc in the supernatant (Figures 3.6A and B, lanes 5 and 6).

Furthermore, to ensure that Yju2p is not responsible for the displacement of SF3a and SF3b, Yju2p was metabolically depleted from cells prior to temperature shift,

harvest, affinity purification and treatment. The above strategy was used for analysis of the Yju2p depleted spliceosome containing Hsh155p-myc. Yju2p, which has been proposed to act downstream from Prp2p in the splicing cycle, is not required for the displacement of SF3b, as Hsh155p is found in the supernatant (Figure3.6C, lanes 5 and 6), indicating it is released upon the action of Prp2p even in the absence of Yju2p.

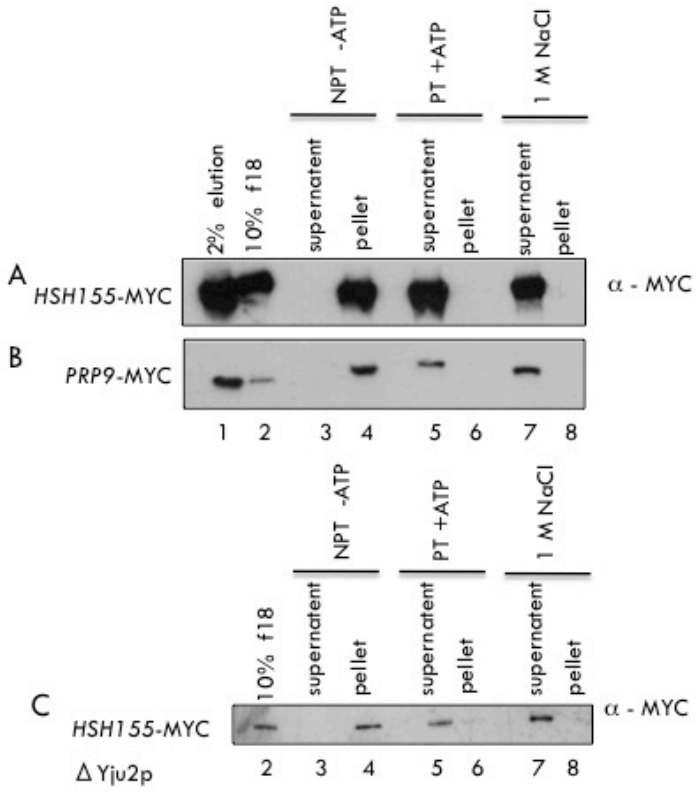


Figure 3.6 SF3 complexes are released from the spliceosome upon Prp2p function.

Affinity-purified spliceosomes were pelleted through glycerol cushions following, treatment at NPT in the absence of ATP, treatment at PT in the presence of ATP or with 1 M NaCl. Total proteins from supernatant and pellet fractions were Western blotting for the MYC tag. The presence of a band in supernatant of the sample treated at PT with ATP indicates that Prp9p (A, lane 5) and Hsh155p (B, lane 5) have been displaced upon Prp2p function. The same experiment was performed in cells depleted of Yju2p. The presence of Hsh155p in the supernatant of the sample treated at PT in the presence of ATP (C, lane 5) supports a role for Prp2p in displacement of SF3 from the spliceosome. Controls include 2% of the total affinity elution and the 10% of peak fraction 18 (f18, lane 2) of the proportion of the particle left untreated as a control.

## 3.4 DISCUSSION

The pre-mRNA splicing machinery and the mechanism of splicing have been extensively studied by both genetic and biochemical means for many years. Although the basic chemical mechanism of the two transesterification reactions of pre-mRNA splicing was proposed 25 years ago (Domdey *et al*., 1984; Padgett *et al*., 1984; Ruskin *et al*., 1984; Will *et al*., 1993), the specific molecular means by which the chemical steps of splicing are authorized and promoted remain unknown. A general mechanism for the function of the DEAH-box protein Prp16p was first proposed by Guthrie and colleagues (Burgess *et al*., 1990). Work by Query and Konarska (Query and Konarska, 2004) has further shaped this model for Prp16p, U6 and Prp8p in the first to second step transition; however, specific functions for both Prp2p and Prp16p in the first and second step, respectively, have best been described as enigmatic rearrangements.

This work illustrates the mechanism for initiating the attack of the branchpoint adenosine on the 5' splice site involves the removal of SF3 from the spliceosome at the time of Prp2p action, thus unsheathing the pre-mRNA brachpoint/U2 snRNA duplex. This data evidences a model in which the U2-associated SF3a and SF3b protein complexes sequester this region and prevent its premature reactivity (see Illustration 3.2).

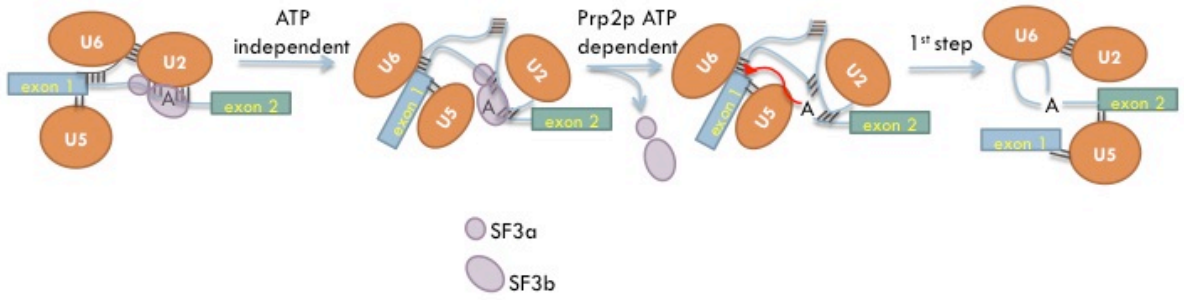


Illustration 3.2 The model for the activation of the first chemical step of pre-mRNA splicing includes an ATP independent conformational rearrangement and the ATP-dependent removal of SF3a and SF3b.

The SF3b complex resides on the branchpoint/U2 snRNA duplex, while SF3a resides just upstream of the branchpoint adenosine during spliceosome assembly, and remains until the ATP-dependent action of Prp2p. Following interaction of Prp2p with the spliceosome, an ATP-independent rearrangement concomitant with a structural rearrangement occurs. Upon ATP hydrolysis by Prp2p, the SF3a and SF3b components are displaced from the pre-mRNA, uncovering the branchpoint adenosine hydroxyl, which is now properly positioned for in-line attack of the 5' splice site. This transesterification releases exon 1 and the lariat intermediate, formed by the 2'-5' linkage between the branchpoint and the 5' splice site.

As described in Chapter 2, compositional analysis of first step- and second step-arrested spliceosomes identified the U2 snRNP subcomplexes, SF3a and SF3b, prior to the first step of splicing (Table 2.4), but indicated the absence of these factors after the first step of splicing (Table 2.5), suggesting that these factors are displaced sometime between the Prp2p action and Prp16p function. Functionally testing this on *in vivo*-purified first step-arrested spliceosomes demonstrates the release of SF3a and SF3b upon ATP treatment, and the removal of these factors is preceded by an ATP-independent spliceosomal rearrangement.

The ATP-independent structural rearrangement occurs upon raising the *prp2*-isolated spliceosome to permissive temperature. Since the arrest of the spliceosome was dependent upon a shift to non-permissive temperatures, the conformational change, as measured by sedimentation through a glycerol gradient and migration in a native gel (Figures 3.2 and 3.4), is proposed to be a result of Prp2p function. This change in the spliceosome may result from a repositioning of a snRNP, the pre-mRNA, or both to favor the arrangement of the reactive sites for the first step (Konarska *et al*., 2006). Upon ATP treatment, SF3a and SF3b are quantitatively removed from spliceosomes (Figures 3.3, and 3.6), and pre-mRNA is chased into the first step products (Figure 3.5) demonstrating that the *prp2*-arrested spliceosomes are functional and are not defective, off-pathway intermediates. The spliceosomes purified in this work were prepared under moderately stringent conditions, and corresponding mass spectrometry data indicates the presence of all expected first step factors and the absence of most second step-specific factors (Chapter 2, Table 2.4), suggesting that the particles purified contain all components sufficient to proceed through the first step of splicing and not the second step of splicing.

Several DExH/D-box proteins have been reported to have both ATP-dependent and ATP-independent functions, including Prp5p during spliceosome assembly (Newnham and Query, 2001; Perriman *et al*., 2003) and Prp22p, which has an ATP-independent role in the second step of splicing (Schwer and Gross, 1998) prior to its ATP-dependent role in displacing the spliceosome from the spliced mRNA (Company *et al*., 1991; Schwer, 2008). Similarly, we have shown a parallel ATP-independent rearrangement at the time of Prp2p function, and defined an ATP-dependent role for Prp2p in initiating the first step of pre-mRNA splicing. Although we cannot yet say definitively that Prp2p is required for the ATP-independent step, the inability to proceed through this step using a *prp2* mutant at non-permissive temperatures strongly suggests that Prp2p plays a role.

The mass spectrometry results presented in Chapter 2 also reveal that the first step factors, Yju2p and Cwc25p are not found in *in vivo* purified first step-arrested spliceosomes. Previous work from the Cheng laboratory and the Lührmann laboratory on *in vitro* transcribed pre-mRNA propose that these two factors act prior to the first step (Chiu *et al*., 2009; Liu *et al*., 2007; Warkocki *et al*., 2009). Our data demonstrate that *in vivo*-assembled spliceosomes do not require Yju2p for the first step, at least for the pre-mRNA tested (Figure 3.5), and that Prp2p-initiated removal of SF3a and SF3b is not dependent upon these factors. However, our data cannot rule out that Yju2p joins after Prp2p and prior to the first step. One possibility for this difference is that Yju2p may be required for the splicing of only a subset of introns in yeast. This seems to be a probable answer to the discrepancy between *in vivo* and *in vitro* data, and indeed, intron microarray analysis has revealed vast differences in splicing efficiency between different splicing factor mutants (Pleiss *et al*., 2007). Another possibility is that the difference between our *in vivo*-assembled spliceosomes and those formed *in vitro* using Yju2p-

depleted extracts reflect pathway-specific differences resulting from transcription-coupled pre-mRNA processing versus those assembled *in vitro*. We note that SF3a and SF3b displacement is quantitative under these conditions, independent of the presence of Yju2p in the spliceosomes.

Structural analysis of the human SF3b complex demonstrated that the SF3b14a protein, which crosslinks to the branchpoint adenosine (MacMillan *et al*., 1994) and the U2 snRNA (Dybkov *et al*., 2006), resides in a cleft through which the pre-mRNA has been predicted to thread (Golas *et al*., 2005). Although yeast SF3b does not contain an ortholog of SF3b14a, the high degree of conservation of the other SF3b subunits suggests that the U2 snRNA/intron branchpoint duplex is sequestered within SF3b in a protective manner.

Others have shown that early in spliceosome assembly, the branchpoint adenosine is arranged in such a way as to expose the reactive 2'-OH upon base pairing to the U2 snRNA (Newby and Greenbaum, 2002). The importance of restraining this highly reactive nucleophile was highlighted in work showing that, in the presence of a defective 5' splice site, the branchpoint can attack the U2 snRNA itself (Smith *et al*., 2007). In our model, SF3b provides a temporal steric barrier prior to the activation of the first step of splicing (Illustration 3.2). Physically obscuring the branchpoint region at the time of base pairing with the U2-associated SF3b complex allows the numerous rearrangements required to proceed through spliceosome assembly to occur without premature branchpoint reactivity. When the spliceosome achieves the proper conformation, an ATP-independent rearrangement occurs. Although we do not have direct evidence for the participation of Prp2p in the ATP-independent event, the block induced by the prp2G551Np at non-permissive temperature suggests that Prp2p plays a role in this process. Prp2p then induces an ATP-dependent change that destabilizes both SF3b from

the branchpoint region, and SF3a from the region upstream of the branchpoint. This unleashes the branchpoint 2'-OH, which is now positioned for the proper in-line attack at the 5'splice site to free the 5' exon and form the lariat intermediate. With this work a mechanistic role is described to Prp2p. In addition, this work provides a system by which further functional testing may elucidate other mechanistic consequences of Prp2p function.

# Chapter 4: Prp24p is not essential for U4/U6 annealing in the presence of high levels of U6 or for activation of the spliceosome.

## 4.1 snRNP RECYCLING AND REASSEMBLY

Following mRNA release by Prp22p, the spliceosome must be disassembled so the individual snRNPs can be recycled for further rounds of splicing. Prp43p and Brr2p are responsible for disassembly of the post-splicing spliceosome in an energy dependent manner (Martin *et al*., 2002; Small *et al*., 2006). Upon release from the spliceosome, U2, U5, and U6 undergo snRNP-recycling events, which prepare them for reentry into the splicing cycle.

### 4.1.1 U6 snRNA

Though all 5 snRNAs are essential for viability and for splicing in yeast, U6 has been regarded as the central snRNA in the splicing cycle and is the most conserved snRNA, with close ~75% identity to its human counterpart (Brow and Guthrie, 1988). The U6 snRNA differs significantly in its biogenesis from the other spliceosomal snRNAs. U6 is transcribed by RNA Pol III, while the other spliceosomal snRNAs are transcribed from RNA Pol II (Kunkel *et al*., 1986). Perhaps as a result of being the only snRNA transcribed by RNA Pol III, the 5' end of U6 is capped with a γ-monomethyl phosphate, as opposed to the 5'-methylguanosine cap that protects the 5' ends of the other snRNAs (Singh and Reddy, 1989).

The U6 snRNA plays a major role in each of the assembly, splicing, and post-splicing events. As the spliceosome must adopt conformations that favor the chemical reactions of splicing to take place, U6 must also be reconfigured within the spliceosome. Thus, U6 snRNA is highly dynamic during the splicing cycle, changing conformations,

possibly even triggering the splicing cycle to continue. Chemical probing experiments have traced the secondary structure of the "naked" U6 snRNA, U6 within its mono-snRNP, and its structure within the U4/U6 di-snRNP (Jandrositz and Guthrie, 1995). Though these structures have provided much insight into the secondary structure of the U6 mono-snRNP, the dynamic nature of U6 snRNA has made the subject a source of further analysis and debate (Dunn and Rader, 2010; Fortner *et al*., 1994; Karaduman *et al*., 2006; Vidaver *et al*., 1999).

### 4.1.2 U6 snRNP

The recycling pathway for the U6 snRNP includes the binding of snRNP components in addition to at least one structural rearrangement, which is thought to be facilitated by Prp24p. Prp24p is an RNA binding protein, which contains four consensus RNA recognition motifs (RRMs), and is found only to co-immunoprecipitate with U6 snRNA, making it part of the U6 snRNP (Shannon and Guthrie, 1991). Unlike the U1, U2, U4, and U5 snRNAs, which are bound by the Sm complex, the 3' end of the U6 snRNA interacts with a heptameric ring of LSm (Like Sm) proteins. All together the U6 mono-snRNP contains eight proteins: Prp24p and the seven LSm proteins.

Interestingly, the binding of the LSm proteins to U6 does not seem to affect the secondary structure of U6 to a great degree (Fortner *et al*., 1994; Jandrositz and Guthrie, 1995); however, interactions between Prp24p and the LSm ring have been shown to be critical in the formation of the U6 snRNP and further downstream events (Rader and Guthrie, 2002). In addition, recent NMR studies have demonstrated the binding of RRM2 of Prp24p to a short stretch of nucleotides within the stem of the U6 snRNA (Martin-Tumasz *et al*., 2010). This binding is thought to recognize the RNA while the stem is

breathing and position RRM1 of Prp24p against U6 to "capture" the upstream sequence of nucleotides, leaving U6 nucleotides unprotected, and available to anneal with U4.

### 4.1.3 U4/U6 annealing

Sequence analysis of U4 and U6 showed extensive complementarity between the two snRNAs, suggesting that they may form two intermolecular stems (Brow and Guthrie, 1988). The duplex was found to have a uniquely high melting temperature of 53°C (Brow and Guthrie, 1988; Siliciano *et al*., 1987). This extensive intermolecular base pairing was further characterized as dependent upon the action of the U6 snRNP component, Prp24p (Jandrositz and Guthrie, 1995; Raghunathan and Guthrie, 1998a; Shannon and Guthrie, 1991). As a result of this, Prp24p has for many years been the subject of structural analysis aimed at understanding how exactly Prp24p facilitates this annealing event (Bae *et al*., 2007; Kwan and Brow, 2005; Martin-Tumasz and Butcher, 2009; Martin-Tumasz *et al*., 2010). Though it is still unclear exactly how Prp24p mediates the interaction between U4 and U6, it is known that upon binding, Prp24p alters the secondary structure of U6, likely making U4 annealing sites available.

Previous reports have established that Prp24p is not a stable component of the U4/U6 annealed di-snRNP (Gottschalk *et al*., 2001b; Stevens and Abelson, 1999); however, a cold-sensitive point mutation in the U4 snRNA arrests a particle containing U4, U6, and Prp24p (Shannon and Guthrie, 1991). Within this stalled particle the melting temperature of the U4/U6 interaction is ~37°C, or 16°C lower than the wildtype melting temperature, indicating that U4 and U6 are not stably base paired in the presence of Prp24p (Jandrositz and Guthrie, 1995; Shannon and Guthrie, 1991) (Lardelli and Stevens unpublished).

### 4.1.4 Prp24p as an antagonist to the U4/U6 interaction

Genetic approaches have been used to further characterize the mechanism of function of Prp24p in respect to di-snRNP. Evidence from these studies suggest that Prp24p acts as an antagonist of the U4/U6 base pairing interaction prior to or immediately upon full annealing between the snRNAs (Ghetti *et al*., 1995; Vidaver *et al*., 1999). This had led several laboratories to propose a role for Prp24p in U4 unwinding during spliceosome activation (Ghetti *et al*., 1995; Strauss and Guthrie, 1994; Vidaver *et al*., 1999). As U4 is unwound from the spliceosome, the U6 ISL is reformed, in addition to several short duplexes between U6 and U2 (see Illustration 1.4). Specifically, the Brow laboratory has characterized a cold-sensitive version of U6 in which the ISL of U6 is hyperstabilized, preventing melting of the region important for binding U4 during U4/U6 annealing (Vidaver *et al*., 1999), resulting in severely reduced levels of the U4/U6 base pairing interaction. Suppressors of this annealing defect have been mapped to the Prp24p; however, these Prp24p mutants fail to suppress the cold-sensitive phenotype, which points toward an additional role for Prp24p. Because the ISL seems to be stabilized by Prp24p, it is possible that Prp24p facilitates the stabilization of the reformed U6 ISL during activation, resulting destabilization of U4 snRNA (Vidaver *et al*., 1999). While strong genetic evidence exists for Prp24p function in activation, biochemical data has yet to support this hypothesis.

### 4.1.5 Other di-snRNP components

The levels of U6 in a yeast cell are roughly five times higher than the levels of U4 within the cell (Li and Brow, 1993). Because of this and the energetically favorable

U4/U6 base paired conformation, levels of free U4 snRNP are undetectable in wildtype yeast cells. Thus it has been challenging to determine the protein contribution from U4 snRNP to di-snRNP. Several factors have been characterized as di-snRNP proteins, but the exact timing of their entrance into the yeast di-snRNP remains to be determined. These include Snu13p, Prp4p, Prp31p, and Prp3p (Anthony *et al*., 1997; Banroques and Abelson, 1989; Weidenhammer *et al*., 1997). Each of these components has also been characterized among the tri-snRNP components.

Snu13p was initially discovered as an essential tri-snRNP component (Gottschalk *et al*., 1999; Stevens and Abelson, 1999) required for proper splicing. Soon after, Snu13p was also shown to play a role in ribosome biogenesis as a component of the box C/D snoRNA complex, which is known to 2' *O*-methylate rRNA during processing (Watkins *et al*., 2000). Snu13p shows remarkable conservation from *S. cerevisiae* to humans, and the human homolog, 15.5K has been shown to bind a kink-turn in the U4 snRNA intramolecular stem loop of U4 in the U4/U6 snRNP and then recruit the human counterpart of Prp31p as well as the heterodimer of hPrp3p and hPrp4p (Nottrott *et al*., 1999). A similar process is suggested to take place in yeast, where Snu13p binding to U4 snRNA precedes the entry of other di-snRNP factors, serving to recruit them to di-snRNP (Oruganti *et al*., 2005).

While *in vivo* data from human cell culture indicates that the human homolog of Prp31p forms an essential interaction with the U5 snRNP factor, Prp6p, in the formation of tri-snRNP (Schaffert *et al*., 2004), studies in yeast suggest that Prp31p serves as a stabilization factor during tri-snRNP integration into the spliceosome and may actually exist both in the spliceosome and in tri-snRNP (Weidenhammer *et al*., 1997). In both systems, Prp31p is essential for splicing, like the other di-snRNP components. Prp31p

requires Snu13p to assemble into di-snRNP; however, it is not known if Prp3p and Prp4p are also required for this interaction.

Prp4p was also discovered in *S. cerevisiae* as a stable component of both di-snRNP and tri-snRNP (Banroques and Abelson, 1989; Bjorn *et al*., 1989). It was found that under low salt conditions Prp4p immunoprecipitated U4/U6•U5 and under even lower salt, B complex spliceosomes. Similar to Prp4p, Prp3p was discovered as an essential component of both di-snRNP and tri-snRNP. Conditional mutants of Prp3p show a severe assembly defect of U4/U6•U5 into spliceosomes with no effect on levels of U4/U6, suggesting that Prp3p may be required for stabilization of U4/U6 and integration of U4/U6•U5 into spliceosomes (Anthony *et al*., 1997)

Prp3p and Prp4p have been shown to form a direct and stable interaction prior to joining the U4/U6 snRNP (Ayadi *et al*., 1998). The two proteins have significant genetic interactions and extensive physical interactions. In addition, Prp4p belongs to the WD-protein family, which is characterized by 4 to 10 WD repeats. This structural motif forms an interface for protein-protein interactions (Neer *et al*., 1994), which in the case of Prp4p, could serve to form a scaffolding surface for the assembly of tri-snRNP, or more simply, could serve as an interface for interaction with Prp3p (Ayadi *et al*., 1998). In addition, the 5' end of the U4 snRNA has been shown to interact with Prp4p, and this interaction plays a critical role in incorporation of U4/U6 into higher order snRNP complexes (Bordonne *et al*., 1990; Xu *et al*., 1990).

Taken together, the Lührmann laboratory has proposed a hierarchical assembly pathway for the di-snRNP components in humans (Nottrott *et al*., 2002), beginning with binding of 15.5K (Snu13p) to the 5' stem loop of the U4 snRNA in the U4/U6 complex. Following this binding event, other components including Prp31p, Prp3p and Prp4p can assemble and serve to stabilize the di-snRNP while U4/U6 joins U5 to form tri-snRNP.

Note that SART3, the human ortholog of Prp24p is not included in this schematic, as it has likely exited the complex.

Finally each of these di-snRNP/tri-snRNP components is also incorporated into spliceosomes (C complex). However, none of the factors is found in complexes succeeding the activated spliceosome (B* complex), indicating they have been destabilized at or around the time of U4 release and spliceosome activation. It has been suggested by Reinhard Lührmann and colleagues that the human homologues of Prp3p and Prp4p, along with another protein that has no known ortholog in yeast, act in stabilizing stem II of the U4/U6 interaction while stem I is reversibly unwound. Furthermore, they also suggest that this could be a mechanism of proofreading (Nottrott *et al*., 2002).

### 4.1.6 Tri-snRNP formation

Following maturation, the U4/U6 snRNP can then interact with the U5 snRNP to form U4/U6•U5 snRNP. The joining between the two snRNPs is based on protein-protein interactions between U4/U6 snRNP components and U5 snRNP components. Specifically, human Prp31p has been shown to play an important role in contacting the U5 snRNP factor; Prp6p, however other contacts are also likely responsible for this salt-stable association.

The role of Prp24p has been well defined as essential for both splicing and cell viability; however, it remains unclear when and how Prp24p is released from the U6 snRNA. While structural analyses and immunoprecipitations suggest that Prp24p binds only U6, it is possible that a U4/U6•Prp24p complex exists in a transient state prior to the subsequent assembly events. Furthermore, the temporal association of the U4/U6 snRNP

components into the snRNP remains elusive. This study serves to gain further knowledge on assembly pathway in yeast spliceosomes beginning with the assembly of the U4/U6 snRNP and the following U4/U6•U5 formation.

## 4.2 MATERIALS AND METHODS

### 4.2.1 Plasmids

| Plasmid | Purpose | Marker |
|---|---|---|
| pRS415 | Control/serial dilutions | *LEU2* |
| pRS425 | Control/serial dilutions | *LEU2* |
| pRS415-*SNR6* | Control/serial dilutions | *LEU2* |
| pRS425-*SNR6* | Experimental/serial dilutions | *LEU2* |
| pRS425-*SNR14* | Experimental/serial dilutions | *LEU2* |
| yCP50-*SNR14* | Strain creation | *URA3* |
| pSE-*snr14*-G14C | Cold-sensitive | *HIS3* |
| pBS1539 | TAP tagging | *URA3* |
| pFA6a-His3MX6-PGAL-3HA | N-terminal HA tag, galactose inducible promoter | *HIS3* |

Table 4.1 Plasmids used in Chapter 4 for tagging or serial dilutions.

### 4.2.2 Yeast strains

| Strain | Mating type | Genotype |
|---|---|---|
| SS330 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS416-*PRP24* |
| SS1100 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS416-*PRP24*, pRS425 |
| SS1101 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS416-*PRP24*, pRS425-*SNR6* |
| SS1102 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS416-*PRP24*, pRS415-*SNR6* |
| SS1106 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS416-*PRP24*, pRS425-*SNR14* |
| SS1103 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS425-*SNR6* |
| SS1104 | a | *leu2Δ1, his3Δ1, ura3Δ0, prp24::KAN*, pRS425-*SNR6*, *LSM8TAP$_a$::URA3* |
| SS419 | α | *leu2Δ0, his3Δ200, ura3Δ0, met15Δ0, trp1Δ63 LSM8TAP$_A$::URA3* |
| SS1394 | α | *leu2Δ0, his3Δ200, ura3Δ0, met15Δ0, trp1Δ63 LSM8TAP$_A$::URA3 TRP::GAL-HA-PRP24* |
| SS1160 | a | *trp1, his3, ura3, ade2, lys2, snr14::TRP1*, YCp50-*SNR14 (URA)* (Shannon and Guthrie, 1991) |
| SS1161 | a | *trp1, his3, ura3, ade2, lys2, snr14::TRP1*, pSE-*snr14-G14C* (Shannon and Guthrie, 1991) |
| SS1162 | a | *trp1, his3, ura3, ade2, lys2, snr14::TRP1*, pSE-*snr14-G14C*, *PRP3TAP::URA3* |
| SS1163 | a | *trp1, his3, ura3, ade2, lys2, snr14::TRP1*, pSE-*snr14-G14C*, *PRP24TAP::URA3* |
| SS1167 | a | *leu2Δ0, his3Δ200, ura3Δ0, met15Δ0, trp1Δ63, PRP3TAP::URA3* |

Table 4.2 Yeast strains used in Chapter 4.

### 4.2.3 Oligos

| Oligo name | Sequence |
|---|---|
| PRP24TAPA | 5’AACAAGAGCAGATGTCCAACGACGATTTTCGCAAGATGTTTCTAGGTGAGTCCATGGAAAAGAGAAG3’ |
| PRP24TAPB | 5’ATGTGAACGTTTGGAAAGAAACTTGGGTTTATATTCACTGCTAAAATGACTACGACTCACTATAGGG3’ |
| PRP3TAPA | 5’ACGTACGCTGGGTCAGTTTGATTCAGAGCATTTTTATTCACCTGTTCAAACGTCCATGGAAAAGAGAAG3’ |
| PRP3TAPB | 5’TATCGGTGTCTACAAAATGATACGCTTTGTTTCATATTAAATATTATTTTACGACTCACTATAGGG3’ |
| LSm8TAPA | 5’TCGAAAATGAGCATGTAATATGGGAAAAAGTGTACGAATCAAAGACAAAATCCATGGAAAAGAGAAG3’ |
| LSm8TAPB | 5’GTTAATGCTTAAATTTATTGTATGATTTATATACTTCTATACATGGTATTTACGACTCACTATAGGG3’ |
| PRP24F4 | 5’TTTACTACAACATTGAAAACCCTTTATATTTATTATATGAGAATTCGAGCTCGTTTAAAC3’ |
| PRP24R3 | 5’GTCGTTTTGAATCTGGTCTAGCGTGATGTCCATACTCCATGCACTGAGCAGCGTAATCTG3’ |
| U1 probe | 5’GAATGGAAACGTCAGCAAACAC3’ |
| U2 probe | 5’AAGAACAGATACTACACTTGA3’ |
| U4 probe | 5’ACCATGAGGAGACGGTCTGG3’ |
| U5 probe | 5’ATGTTCGTTATAAGTTCTATAGGC’3 |
| U6 probe | 5’AGGGGAACTGCTGATC3’ |
| U3 probe | 5’TGATCCTACTGTAGAATGTGTTAGTCAAAA3’ |

Table 4.3 Oligos used in Chapter 4 for tagging or Northern blotting.

### 4.2.4 Strain construction

Yeast strains listed in Table 4.3 were created by the lithium acetate transformation method (Gietz and Woods, 2002). Plasmids were transformed into to parent strains and selection was maintained. Strain SS1161 was created by transforming with and selecting for pSE-*snr14-G14C*. To screen for colonies that lost the wildtype copy of *SNR14*, cells were dilution plated onto minimal media (lacking HIS) and then plated on 5'FOA, which kills cells containing the product of the *URA3* gene.

C-terminally tagged strains were created through use of a PCR-generated fragment with arms of homology to the region in which the tag was to be knocked in, followed by a selectable marker, *URA3*. Colonies that grew on synthetic media were screened for presence of a functional TAP tag by western blotting with the peroxidase anti-peroxidase antibody (Rockland) (see Chapter 2, section 2.2.4). N-terminal HA tags under a galactose inducible promoter were created through use of the same PCR-based system, using arms of homology upstream of the gene being tagged, and transformants were tested for presence of a function HA tag, as described in Chapter 3, section 3.2.4.

### 4.2.5 Serial streaking and dilutions

Following transformation of plasmids into the parent strain, transformants were streaked out on SD-LEU for the strains containing one of the following vectors: pRS425, pRS425-*SNR14*, pRS425-*SNR6*. Cells were restreaked on SD-LEU two times, and then streaked onto SD-LEU, SD-URA or SD-LEU + 5'FOA at 31°C. For serial dilutions, single colonies were grown in liquid minimal medium. OD's were normalized to one optical density (OD) unit and dilutions were made. 3µl of each dilution was plated on SD-LEU, SD-URA, or SD-LEU + 5'FOA at 31°C.

### 4.2.6 Native total RNA extraction

Native RNA was extracted from strains in which base pairing was to be tested. Native RNA was harvested according to the guanidium thiocyanate (GTC) method, as described by Wise *et al*. (Wise *et al*., 1983). Briefly, cells were harvested and washed once in diethylpyrocarbonate (DEPC) treated water. One cell volume of RNase free beads was added to the cells followed by one cell-volume of GTC. The mixture was vortexed for 45 seconds followed by a 15 second ice incubation. This was repeated 4 times. Ten cell volumes of DEPC water was added, followed by ten cell volumes of DNA phenol. The mixture was vortexed two times with a one-minute incubation on ice in between vortexing. The extraction was then centrifuged at 4.4 x g for 5 minutes at room temperature. The aqueous layer was removed and extracted 3 more times. To the final extracted aqueous layer, 10% volume of 3 M sodium acetate, pH 5.3 and 2.5 volumes of ice-cold ethanol were added. RNA was precipitated overnight at -20°C.

### 4.2.7 Non-denaturing gel electrophoresis

Native-extracted RNA samples were collected after precipitation by centrifugation and were then resuspended in DEPC water. Separation was conducted by electrophoresis through a 20 x 16 x 0.75 mm gel containing 9% 19:1 acrylamide: bis-acrylamide, 1xTBE gel. The gel was run in 1xTBE running buffer at 25 mA for 2.5 hours at 4°C. Following electrophoresis, the gel was transferred to nylon membrane for Northern blotting.

### 4.2.8 Northern blotting

Gels were either transferred to nylon membrane (Whatman) at 10V overnight at 4°C or at 0.75 A for 3 hours at 4°C in an electroblotting apparatus (Bio-Rad) in 25 mM sodium phosphate, pH 7.0. RNA was crosslinked to the membrane by exposure to 254 nm UV light for 1 minute. Northern blotting was conducted as described in Chapter 2, section 2.2.8.

### 4.2.9 Psoralen crosslinking of U4/U6

Psoralen crosslinking was performed as previously described (Wassarman, 1993) using the psoralen derivative: aminomethyltrioxsalen (AMT). A 400µl glycerol gradient fraction was split into four equal fractions. Samples were treated as follows:

1. no 365 nm UV light and no AMT
2. no 365 nm UV light and + 20 µg/µl AMT
3. +365 nm light and no AMT
4. +365 nm light and + 20 µg/µl AMT. Samples exposed to UV light were incubated at a distance of ~2 cm from the light for 20 minutes. Samples not exposed to UV light were left on ice. The samples were then phenol/chloroform extracted and precipitated using 2.5 times the volume of ethanol. Samples were separated on a denaturing gel (7% 19:1 acrylamide: bisacrylamide, 1x TBE, 8 M urea) at 25 mA/gel for 1 hour and 45 minutes. The gel was transferred to nylon and Northern blotted with radiolabeled oligonucleotides designed to recognize U6 snRNA.

### 4.2.10 Prp24p depletion

To deplete Prp24p from SS1394, cells were grown in rich media (Hong *et al.*) containing 2% galactose to an OD of 0.5. Cells were harvested by centrifugation at 4.4 x g for 5 minutes at room temperature, and washed with pre-warmed YP containing 2% dextrose, and then centrifuged again. Cells were then resuspended in YPD, and grown for 24 hours, and cells were maintained under an OD of 1.0. Aliquots were taken every 1.5 hours. These aliquots were used for extraction of native RNA as well as protein preparation for Western blot analysis. Western blot analysis was conducted as described in Chapter 3, using the α-HA antibody at a 1:1000 dilution in BLOTTO, and the goat-α-mouse, HRP conjugated secondary antibody.

### 4.2.11 Extract preparation

Whole-cell extract was prepared according to the following protocol. First, cells were grown to an OD ($A_{600}$) of ~1.5. At this point, cells were harvested, washed with ice-cold water, washed with AGK buffer (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 8% glycerol). Cells were then resuspended in 1 mL AGK per 1 g of cells and flash frozen in pellets by dropping into liquid nitrogen. Next cells were disrupted by ball-milling in a Retsch MM300 mixer mill with 50 mL canisters immersed in liquid nitrogen. Briefly, cells were disrupted by shaking at 10 Hz for 5 cycles of 3 minutes each. Canisters were dipped into liquid nitrogen in between each cycle. The pulverate was then thawed, and allowed to extract for 30 minutes by gentle rotation at 4°C in the presence of 5 μM Leupeptin, 3 μM Pepstatin, 0.4 mM PMSF and 1mM DTT. Extract was cleared by centrifugation at 30,000 x g for 30 minutes at 4°C, and then 100,000 x g for 1 hour at

4°C. The middle third of the cleared extract was gently removed following the second centrifugation step, and allowed to dialyze in Buffer D (20% glycerol, 20 mM HEPES, pH 7.9, 50 mM KCl, 0.5 mM DTT, 1.5 mM $MgCl_2$) twice for 2 hours at 4°C. After dialysis, the extract was centrifuged at 20,817 x g for 10 minutes at 4°C. The extract was frozen in liquid nitrogen and stored at -80°C.

### 4.2.12 Native gel analysis of splicing complexes

Electrophoresis was used to separate splicing complexes from 4 μL of extract (as prepared in 4.2.7). Complexes were run on a 20 cm x 16 cm x 1.5 mm gel with a composition of 4% 80:1 acrylamide: bisacrylamide, 1 x TGM buffer (50 mM Tris base, 50 mM glycine, 2 mM $MgCl_2$) at 160V for 6 hours at 4°C with recirculating buffer (Raghunathan and Guthrie, 1998b). The gel was then transferred to nylon or nitrocellulose for Northern blotting or western blotting, respectively.

### 4.2.13 Affinity purification of di-snRNP and tri-snRNP

Cells were harvested, disrupted, and extracted as described in Chapter 2, section 2.2.7. Affinity purifications were carried out as described in Chapter 2, section 2.2.7. Following TEV cleavage, the elution was layered onto a 10-30% glycerol gradient (for tri-snRNP) or a 10-20% glycerol gradient (for di-snRNP) with a composition of 20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.02% NP-40, 150 mM KCl, 1 mM DTT, 0.4mM PMSF. Gradients were centrifuged at 29,000 rpm for 16 hours at 4°C for tri-snRNP separation and 36,000 rpm for 20 hours at 4°C for di-snRNP separation in an SW-41 rotor at 4°C. Fractions were phenol/chloroform extracted, and RNA was precipitated in the presence of 20 μg of glycogen (Roche) and 0.1% 3 M sodium acetate, pH 5.3 with 2.5

times ice-cold ethanol. RNA was analyzed by native and/or denaturing gel electrophoresis as described in section 4.2.7 and Chapter 2, section 2.2.8, respectively.

### 4.2.14 Mass spectrometry analysis

Proteins corresponding to the peak fractions of spliceosomal assembly intermediates were precipitated, pooled, and lyophilized. Protein digest with trypsin was carried out as described in Appendix IV by Michelle Gadush, and liquid chromatography separation and mass spectrometry analysis was performed as described in Appendix IV by Dr. Maria Person in the Protein Microanalysis Facility at University of Texas, Austin.

### 4.2.15 Duplex melting

Glycerol gradient particles were brought to room temperature and incubated with 1% SDS and 1 mg/mL proteinase K for 1.5 hours. Particles were diluted 1:5 in a buffer containing 20 mM HEPES, pH 7.9 and 150 mM KCl. The diluted deproteinized particle was divided into 50 μL aliquots and incubated at the 20, 45, 48, 51, 54, 57, 60, 63, or 66°C for 5 minutes. Reactions were flash frozen in a dry ice ethanol bath. Samples were brought to 20°C and immediately loaded on a 7% 19:1 polyacrylamide, 1x TBE non-denaturing gel, and separated at 15 mA for 5 hours at 4°C, with recirculating 1x TBE running buffer. RNA was transferred to nylon, crosslinked, and U6 snRNA was probed for using radiolabeled oligonucleotides.

## 4.3 RESULTS

### 4.3.1 Overexpression of U6 bypasses the essential factor, Prp24p

Previous studies have shown increased levels of U6 snRNA in conditional Prp24p mutant strains (Vidaver *et al*., 1999), suggesting increasing the wildtype levels of U6 can compensate for a handicapped Prp24p. Additionally, an upregulation of U6 snRNA was noted in the absence the zebrafish ortholog of Prp24p, p110 (Trede *et al*., 2007).

To investigate the nature of the relationship between Prp24p and U6, a *PRP24* shuffle strain was obtained. This strain contains a genomic knockout of *PRP24* covered by a *URA3* vector carrying *PRP24*. Into this strain, the following *LEU2*-marked plasmids were transformed: pRS425 (2μ empty vector), pRS425-*SNR6 or* pRS425-*SNR14*. *SNR6* encodes U6 snRNA, while *SNR14* encodes U4 snRNA. Transformants were dilution plated three times on SD-LEU to allow for loss of the wildtype *PRP24* vector. Following dilution plating, cells were streaked out on either SD-LEU, SD-URA or SD-LEU + 5'FOA at 31°C. 5'FOA is a drug that selectively kills any cells harboring a *URA* gene, therefore, any cells still carrying the wildtype *PRP24* vector will not grow. As shown in the third panel of Figure 4.1A, pRS425-*SNR6* can in fact bypass the need for *PRP24*. To test for a growth defect, the strains containing pRS425 or pRS425-*SNR6* were serial diluted on either SD-LEU, SD-URA or SD-LEU + 5'FOA at 31°C. Serial dilutions confirm that the 2μ vector, overexpressing U6 snRNA, can bypass the need for the essential nature of *PRP24* with only a slight growth defect, while controls including the empty vector cannot grow on 5'FOA because they require the *PRP24* vector for viability. The growth on SD-URA implies that the dilution plating prior to plating on 5'FOA is not sufficient for all cells to lose the wildtype copy of *PRP24*; however, cells growing on

5'FOA were then plated on SD-URA to confirm the loss of *PRP24*. For the remaining part of this chapter this strain is termed *Δprp24sup* for *Δprp24* suppressor.

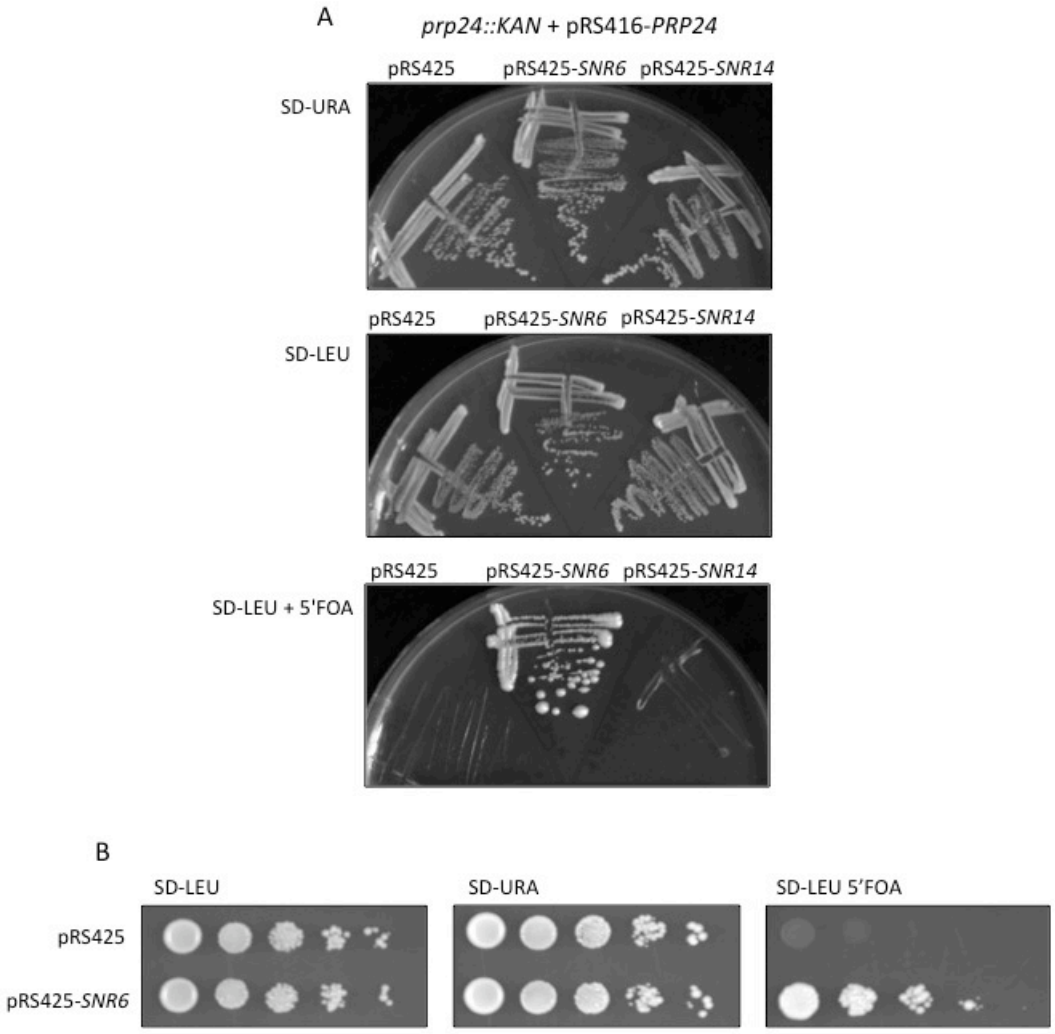


Figure 4.1 High copy U6 snRNA can bypass the need for *PRP24*.

Strains harboring pRS425, pRS425-*SNR6*, or pRS425-*SNR14* were streaked on SD-LEU, SD-URA, or SD-LEU + 5'FOA to test for viability. When streaked (A) or serial diluted (B) on SD-LEU + 5'FOA, cells harboring the *PRP24* gene (which is introduced to the cells on a *URA3* vector), are not viable. However, those cells that lose the *URA3, PRP24* vector will grow. In the presence of high copy U6 (pRS425-*SNR6*), *PRP24* is not essential, and only shows a slight growth defect (B).

### 4.3.2 U4/U6 base pairing analysis by non-denaturing gel electrophoresis

The next step was to determine whether or not U4 and U6 could still form a base pairing interaction in the *Δprp24sup*. Total native RNA was harvested from cells containing wildtype *PRP24* and either *pRS425, pRS415-SNR6,* or *pRS425-SNR6*, and from the *Δprp24sup* strain. Extracted native RNA was then either heated to denature the U4/U6 base pairing or left on ice, and then run on a native gel, transferred to nylon membrane and consecutively probed for U4 and U6 snRNAs to determine the base pairing status between U4 and U6.

As can be seen in the native Northern blots in Figure 4.2, all native samples contain levels of U4/U6 base paired; this result indicates that even in the absence of Prp24p, U4 and U6 can anneal. This suggests that high levels of U6 can indeed bypass the need for Prp24p in facilitating the annealing of U4 and U6, although the levels of U4/U6 is reduced by roughly half that of the wildtype .

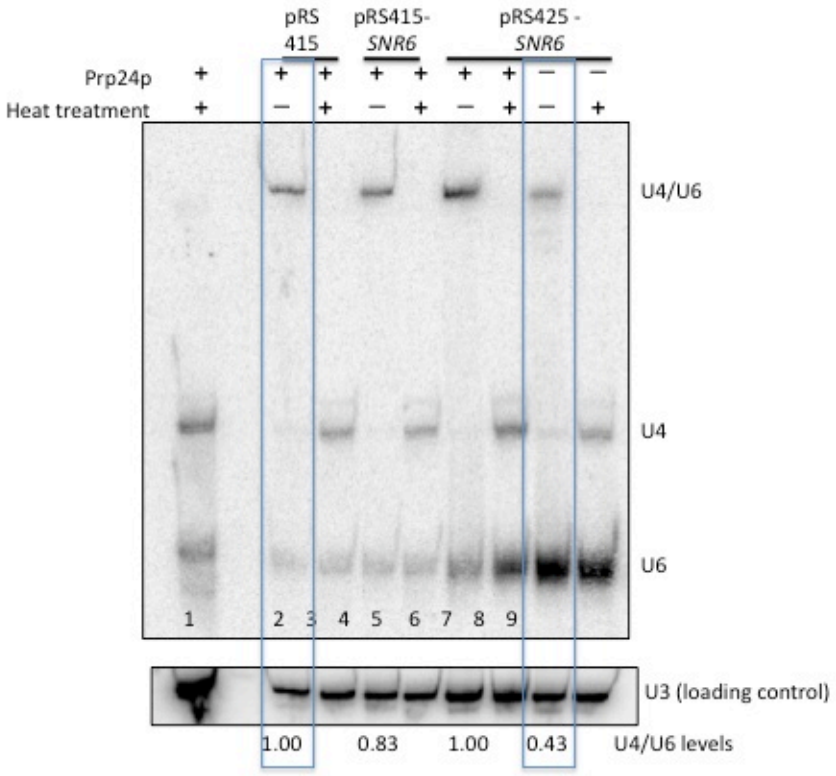


Figure 4.2 U4 and U6 base pair in the absence of Prp24p.

Native total RNA was extracted from wildtype Prp24p strains in the presence of pRS415, pRS415-*SNR6,* or pRS425-*SNR6,* and cells with the genomic disruption of *prp24* containing pRS425-*SNR6*. RNA was separated on a non-denaturing gel, transferred and Northern blotted for U6, U4, and U3, consecutively. The level of U4/U6 in strains lacking Prp24p is decreased by roughly one half that of strains carrying a wildtype copy, as determined by quantitation of U6 in the band representing U4/U6, and normalization to the U3 band.

### 4.3.3 U4/U6 base pairing analysis upon Prp24p depletion

Immunodepletion of Prp24p defined the function of Prp24p in annealing U4 and U6, and in addition showed *in vitro* splicing defects (Raghunathan and Guthrie, 1998b). In light of the results demonstrating that Prp24p is not essential for the base pairing of U4/U6 in strains overexpressing U6 snRNA, it was necessary to confirm that Prp24p is essential for *in vivo* U4/U6 base pairing, and characterize any defects caused by depletion of *PRP24* during growth of cells.

To do this, a yeast strain that could be metabolically depleted of *PRP24* upon shifting carbon sources was created. As shown in Figure 4.3A, cell growth is affected 9 hours after shifting to dextrose. Likewise, levels of Prp24p are not detectable by Western blot for the C-terminal HA tag on Prp24p after 4.5 hours in dextrose (Figure 4.3B). This indicates that upon depletion of Prp24p, cells do not maintain viability, likely due to the lack of recycling of U6 snRNA in the splicing cycle.

Next, the base pairing of U4 and U6 was analyzed in 1.5-hour timepoints by native RNA extraction, followed by non-denaturing gel electrophoresis and Northern blotting for U6, U4, and U3, consecutively. As can be seen in Fig 4.3C, U4/U6 levels begin to drop 1.5 hours after shifting the carbon source to dextrose, which represses the galactose promoter. As U4/U6 levels decrease, there is a concomitant increase in free U4 snRNA levels until 6 hours, at which point levels begin to drop likely due to overall RNA degradation. Interestingly, there is a decrease in levels of U6 snRNA as metabolic depletion proceeds, indicating possible degradation of total cellular levels of U6. This is contrary to what was initially hypothesized - that there would be an increase in expression levels of U6 as levels of Prp24p decreased. However, the interpretation that

U6 levels drop because total cellular RNA is compromised is consistent with drop in levels of free U4.

Finally, the panel probed for the intron containing U3 snRNA (Figure 4.3C) shows an accumulation of unspliced U3 after 4.5 hours of a temperature shift, indicating a splicing defect associated with the Prp24p depletion. This data is consistent with previous reports (Ghetti *et al*., 1995; Raghunathan and Guthrie, 1998b), and reflects an *in vivo* U4/U6 annealing defect, concomitant with a splicing defect upon depletion of the essential Prp24p protein. The *in vivo* demonstration that cell viability decreases upon depletion of Prp24p (Figure 4.3A), in combination with the previous result showing that the functioning of Prp24p in U4/U6 annealing can be bypassed in the presence of high levels of U6, lead to the question of whether or not Prp24p is required for another assembly event in addition to its role in annealing U4 and U6.

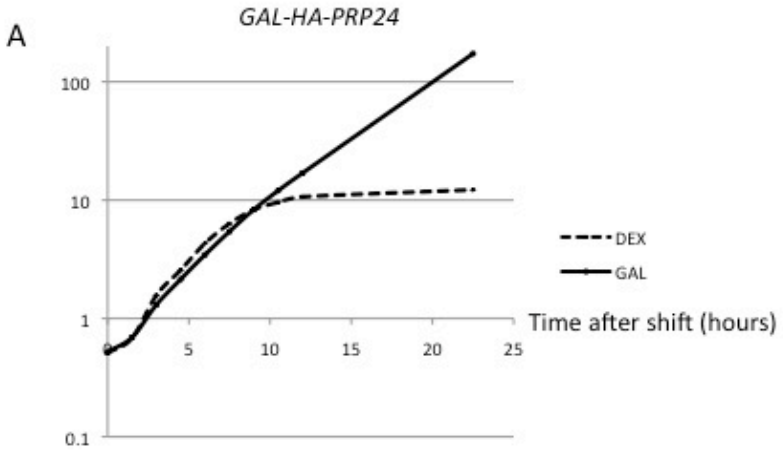


Figure 4.3A *GAL-HA-PRP24* cells are not viable after 9 hours in repressive media.

A strain containing a metabolically depletable version of *PRP24* was shifted to repressive media (dextrose) or maintained in non-repressive media (galactose). The optical density was taken every 1.5 hours and plotted versus time. Following a 9-hour shift to dextrose to deplete *PRP24*, cell growth levels off, indicating the essentiality of *PRP24*.

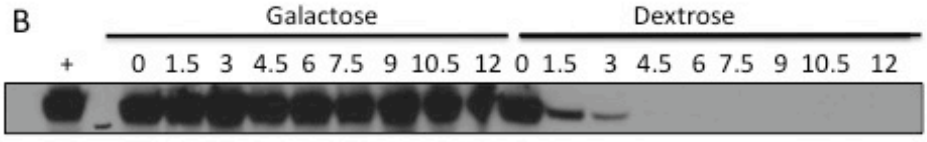


Figure 4.3B Prp24p is not detectable following shifting to repressive media.

Whole-cell extract from each timepoint during Prp24p depletion was separated by SDS-PAGE and subjected to Western blotting for HA-Prp24p. Prp24p is not detectable after shifting to 4.5 hours of growth in dextrose.

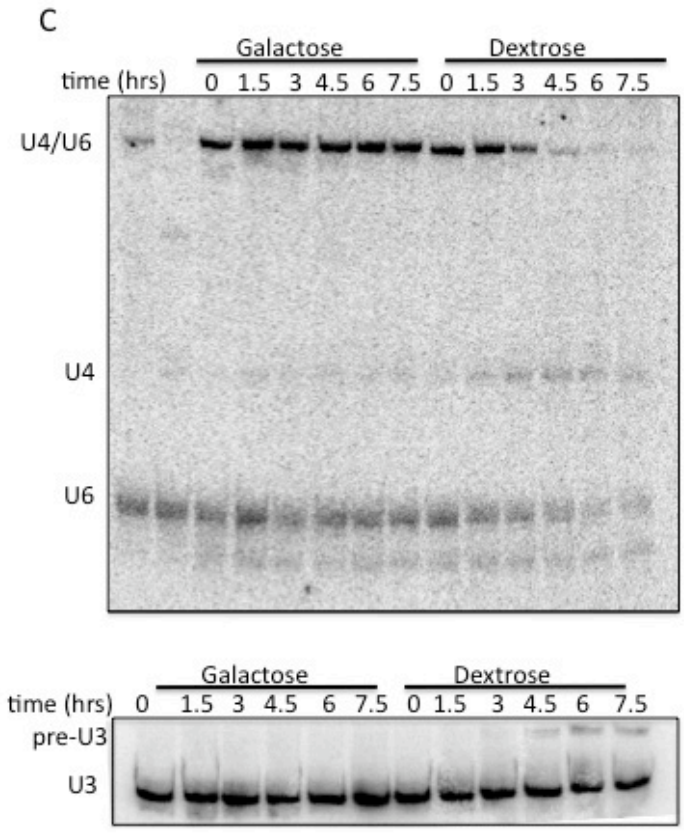


Figure 4.3C. Depletion of Prp24p is concomitant with a decrease in U4/U6 base pairing.

Total native RNA was extracted from each of the timepoints between the sugar shift (T = 0 hrs) and 7.5 hours after the shift, and separated on a native gel. The gel was transferred to nylon for consecutive probing for U6, U4, and U3 snRNAs. U4/U6 levels decrease as levels of free U4 increase. As shown in the bottom panel probed for U3, there is also a splicing defect, which arises as *PRP24* is depleted and U4/U6 levels decrease.

### 4.3.4 Native gel analysis of Prp24p knockout assembly complexes

To further understand and characterize the strain lacking Prp24p, splicing complex conformations were investigated. Nuclear extracts, from cells that had either been grown at 31°C or 16°C were analyzed by native snRNP-resolving gel electrophoresis and Northern blotting for U6 snRNA. The migration of snRNP complexes was analyzed from *Δprp24sup* strain, and wildtype strains harboring no vector, an empty vector (pRS425), or a 2μ vector encoding *SNR6* (pRS425-*SNR6)*.

As seen in Figure 4.4, di-snRNP migrates significantly faster than di-snRNP levels from strains carrying a wildtype copy of *PRP24*. This is interesting considering the finding that U4 and U6 can still base pair in the absence of Prp24p (Figure 4.2). Additionally, tri-snRNP from *Δprp24sup* strains at both 31°C and 16°C migrates noticeably slower than wildtype tri-snRNP. Native gel electrophoresis separates by both size and charge. The slower migration of tri-snRNP resulting from a *Δprp24sup* strain could be due to a change in size or conformation of the complex, which could be accounted for by addition of proteins to the tri-snRNP. Another possibility is that there could be a significant reduction in levels of exposed RNA. Because RNA is highly negative in charge, protection of RNA by protein additions may slow the mobility of the complex.

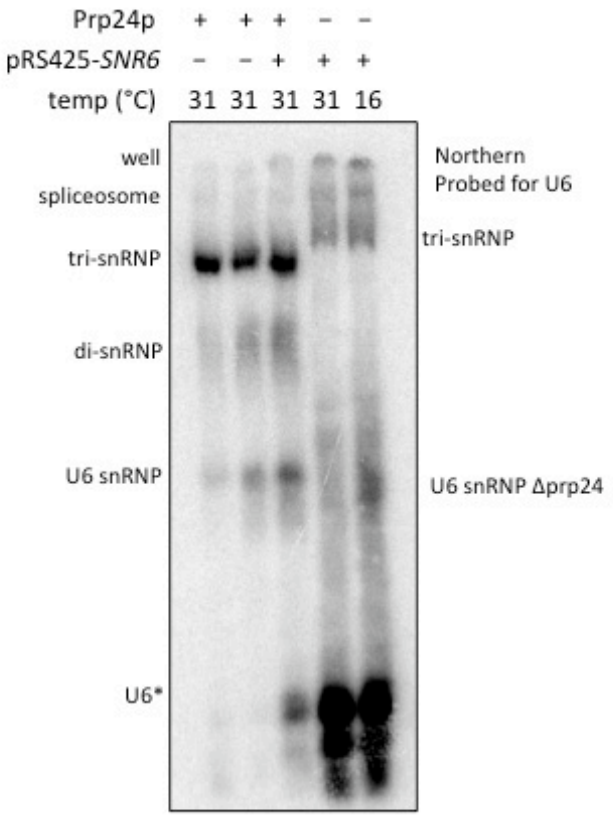


Figure 4.4 Aberrant migrations of di-snRNP and tri-snRNP in *Δprp24sup* extracts indicate that Prp24p may function in events following the annealing of U4 and U6.

Native gel analysis of snRNP complexes from extracts shows that both di-snRNP and tri-snRNP migrate with shifted mobilities when Prp24p is knocked out in the presence of high copy U6, implying that Prp24p, which has been previously described to facilitate the interaction between U4 and U6 (Raghunathan and Guthrie, 1998a), may also be acting in another spliceosome assembly event.

### 4.3.5 Biochemical characterization of *in vivo* purified tri-snRNP from wildtype and *Δprp24sup* strains

Native gel analysis demonstrates that in the absence of Prp24p di-snRNP migrates significantly faster and a conformational defect in tri-snRNP results, indicating that Prp24p may have an additional uncharacterized function in higher order complexes of the spliceosome. To specifically understand the consequences of deleting Prp24p, an affinity purification strategy was used.

#### *4.3.5.1 Tri-snRNP isolation from wildtype and Δprp24 strains*

In order to assess tri-snRNP, LSm8p was C-terminally TAP tagged in the *Δprp24sup*, and Prp3p was C-terminally TAP tagged in a wildtype Prp24p strain. Affinity purifications, as described in Chapter 2, section 2.2.6, were designed specifically to purify tri-snRNP. Following cleavage with TEV protease, the elutions were layered upon a 10-30% glycerol gradient as described in Chapter 2, section 2.2.6, and centrifuged specifically to isolate tri-snRNP. After glycerol gradient fractionation and RNA precipitation, native and denaturing Northern blotting was used to analyze the quality and assess the migration of the spliceosomal intermediates purified. The native northern blot was probed for U6 and U4, consecutively, while the denaturing northern blot was probed for all five snRNAs.

The denaturing Northern blot of spliceosomal intermediates in Figure 4.5A, indicates the presence of purified tri-snRNP in fractions 16 – 20 of the gradient. In addition, U4/U6 snRNP is observed in fractions 8 - 10 of the glycerol gradient. Non-denaturing gel electrophoresis of the identical RNA fractions (Figure 4.5B) shows that U4 and U6 are base paired in both di-snRNP and tri-snRNP. Upon purification and

glycerol gradient sedimentation of tri-snRNP in the strain lacking Prp24p, a similar migration of tri-snRNP in lanes 16-18 is observed (Figure 4.5C); however, the peak does not extend into fraction 20. This could imply there was less starting material, that there was a slight shift in sedimentation, or it may be an indication that there are reduced or unstable levels of tri-snRNP formed from strains lacking Prp24p. Additionally, di-snRNP is sediments in lanes 8-10, while mono-U6 snRNA sediments in lanes 4-6. These results indicate that the LSm complex resides on the U6 snRNA in each of these assembly stages of the spliceosome even in absence of Prp24p. The native northern blot in Figure 4.5D shows that di-snRNP and tri-snRNP contain U4 and U6 mostly in a base paired form. This finding is consistent with previous results (section 4.3.2, Figure 4.2) and demonstrates that not only can U4 and U6 base pair in di-snRNP in the absence of Prp24p, but also that the two snRNAs remain base paired in tri-snRNP.

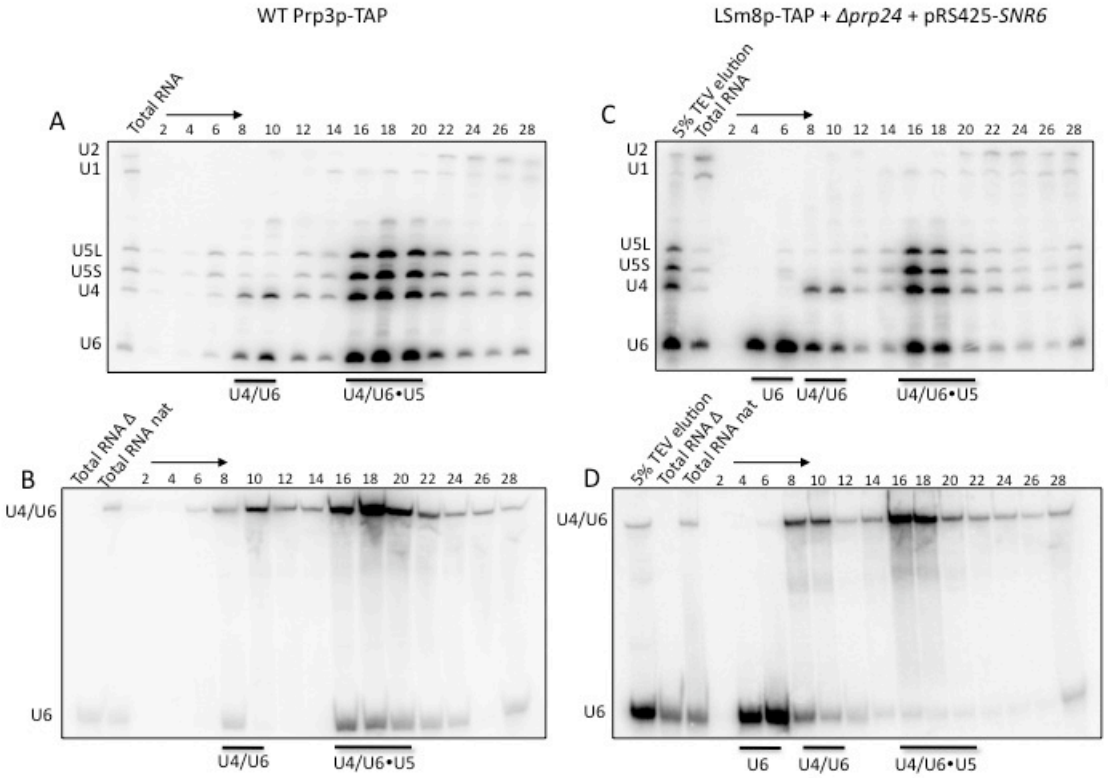


Figure 4.5 Purified tri-snRNP from *Δprp24sup* strain (*right*) shows a similar migration to that purified from wildtype Prp24p (*left*) strains.

Affinity purifications designed to isolate tri-snRNP from both wildtype (A and B) and Prp24p knockout (C and D) strains reveal no significant sedimentation difference between the two resulting tri-snRNP particles. Native gel electrophoresis demonstrates that assembly intermediates from both wildtype (B) and *Δprp24sup* strains (D) contain mostly base paired U6 and strains form base pairing interactions in both di-snRNP and tri-snRNP

#### *4.3.5.2 Sedimentation behavior between tri-snRNP from wildtype and Δprp24sup strains*

U4 snRNA was quantitated across the glycerol gradients from both the wildtype and *Δprp24sup* strains to see if there is a detectable difference in sedimentation behavior between the two tri-snRNPs. As shown in Figure 4.7, there is a slight sedimentation shift in tri-snRNPs produced from a strain lacking Prp24p. This shift is indicates that the tri-snRNP from *Δprp24sup* strains is either of lower molecular weight or of a smaller shape when compared to tri-snRNP from wildtype strains, implying that there may be compositional or conformational consequences to deleting Prp24p, which affect tri-snRNP formation.

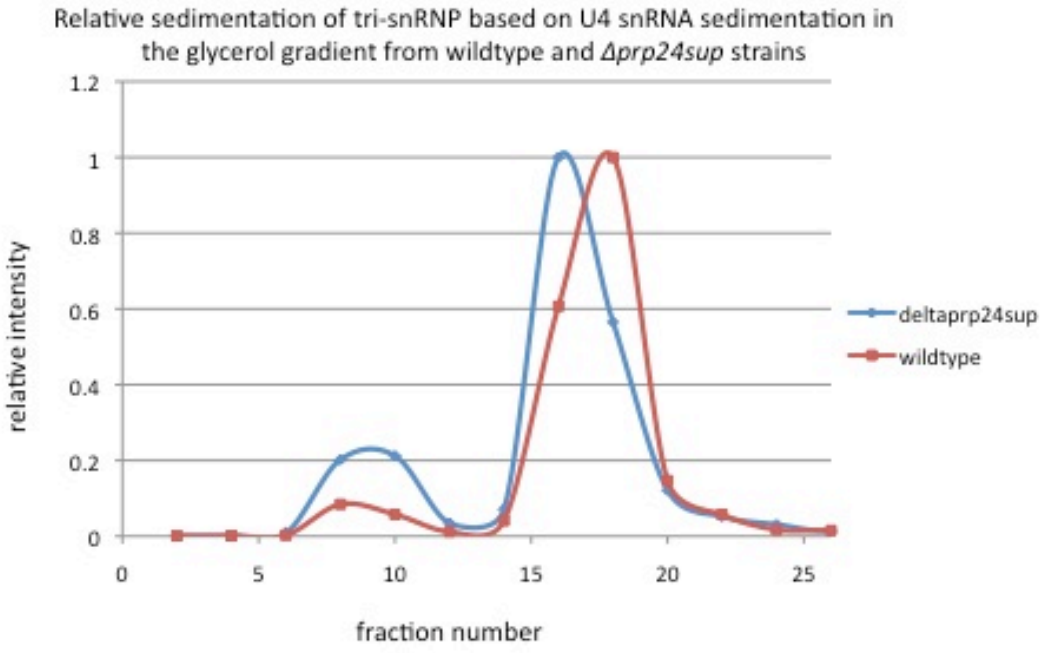


Figure 4.6 Relative sedimentation of tri-snRNP based on U4 snRNA intensity.

U4 snRNA was quantitated across the tri-snRNP glycerol gradients (Figure 4.5A and C). Tri-snRNP produced from *Δprp24sup* strains sediments slightly higher in the glycerol gradient, suggesting either a compositional difference or a slight conformational difference. These results imply that Prp24p may have a role in tri-snRNP formation that has been previously uncharacterized.

#### *4.3.5.3 Compositional comparison between tri-snRNP formed from wildtype and Δprp24sup strains*

The differences observed between wildtype tri-snRNP and *Δprp24sup* tri-snRNP in migration in the native gel (Figure 4.4) and in the sedimentation in glycerol gradients (Figure 4.5), could be due to a compositional variance. To test for this discrepancy between tri-snRNP particles, protein fractions were pooled of tri-snRNP purified from both wildtype and *Δprp24sup* strains and subjected to mass spectrometry analysis.

Table 4.4 shows a comparison between the protein composition of tri-snRNP purified from wildtype and *Δprp24sup* strains. Consistent with glycerol gradient sedimentation, there are very few differences. Both particles contain the heterotrimer of Prp8p, Brr2p, and Snu114p, in addition to the other previously defined tri-snRNP components (Stevens and Abelson, 1999). Interestingly, tri-snRNP from *Δprp24sup* purified particle contains Cbf5p. Cbf5p is a component of the H/ACA box snoRNP, which is known to pseudouridylate the U2 snRNA (Ma *et al*., 2005). Though the snRNP particle under investigation does not contain the U2 snRNA, it is possible that the snRNAs within this tri-snRNP particle are also pseudouridylated. The addition of these this factor likely does not account for the migration shift seen in tri-snRNP of the *Δprp24sup* strain (Figure 4.4). Additional differences noted in the *Δprp24sup* tri-snRNP include the absence of the tri-snRNP-specific factors, Prp38p and Snu23p. While Prp38p has been implicated in activation of the spliceosome (Xie *et al*., 1998), no functional role has been assigned to Snu23p.

| **Protein** | **MW (kDa)** | **Wildtype tri-snRNP (# of peptides IDed)** | ***Δprp24sup* tri-snRNP (# of peptides IDed)** |
|---|---|---|---|
| Prp8p | 279 | 123 | 74 |
| Brr2p | 246 | 73 | 91 |
| Snu114p | 114 | 44 | 54 |
| Prp6p | 104 | 31 | 44 |
| Snu66p | 66 | 19 | 21 |
| Prp3p | 56 | 26 | 29 |
| Prp31p | 56 | 25 | 28 |
| **Cbf5p** | **55** | **-** | **6** |
| Prp4p | 52 | 18 | 15 |
| Spp381p | 34 | 2 | 2 |
| **Prp38p** | **28** | **1** | **–** |
| **Snu23p** | **23** | **3** | **–** |
| SmB1p | 22 | 5 | 6 |
| LSm4p | 21 | 3 | 3 |
| Dib1p | 17 | 3 | 8 |
| SmD1p | 16 | 5 | 5 |
| Snu13p | 13 | 1 | 1 |
| SmD2p | 12 | 6 | 4 |
| LSm8p | 12 | 4 | 6 |
| SmD3p | 11 | 7 | 5 |
| LSm2p | 11 | 2 | 5 |
| LSm5p | 10 | 1 | 4 |
| LSm3p | 10 | 3 | 3 |
| LSm6p | 9 | 5 | 6 |
| SmFp | 9 | 2 | 3 |

Table 4.4 Compositional comparison between tri-snRNPs purified from wildtype and*Δprp24sup* strains reveal no major compositional differences.

Proteins from tri-snRNP purifications (Figure 4.5) were pooled and subjected to mass spectrometry analysis. As shown by bolded text, differences between tri-snRNPs purified from wildtype and *Δprp24sup* strains reveal the absence of Prp38p and Snu23p, and the presence of Cbf5p in tri-snRNP from the *Δprp24sup* strain.

#### *4.3.5.4 U4/U6 melting temperature analysis*

The compositional analysis of the tri-snRNP particles suggest that there is no substantial change in total molecular weight of the tri-snRNP purified from *Δprp24sup* cells. Native gel analysis of intermediates of the splicing cycle demonstrates that there is a defect in tri-snRNP, as implied by the slower migration of tri-snRNP in *Δprp24sup* extract. Since roughly only half of U4/U6 is formed in this strain (Figure 4.2), it could be that there is a deficiency in forming each of the stems of the base pairing interaction between U4 and U6.

To test the stability of the duplex, a glycerol gradient fraction of tri-snRNP purified from both wildtype and *Δprp24sup* cells was proteinase K treated in the presence of SDS, and then divided. Each sample was incubated at designated temperatures ranging from 45°C to 66°C. Wildtype melting temperatures for the U4/U6 duplex are ~53°C (Brow and Guthrie, 1988); however, previously described mutations have lowered this to 40°C (Shannon and Guthrie, 1991). Following temperature incubation, the samples were frozen, then brought to ambient temperatures, and separated by non-denaturing electrophoresis.

Figure 4.7 depicts the melting of the U4/U6 duplex from both wildtype tri-snRNP and tri-snRNP from *Δprp24sup* cells. Both deproteinized U4/U6 duplexes melt at roughly 54°C, indicating that there is nothing substantial interfering with the U4/U6 interaction in tri-snRNPs formed from assembly intermediates lacking Prp24p. This implies that Prp24p is not required for the maintenance of the remarkable base pairing interaction between U4 and U6.

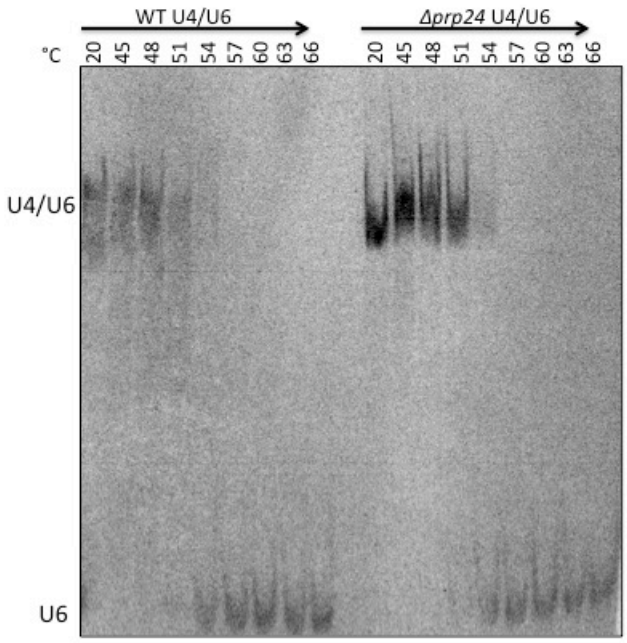


Figure 4.7 Melting of deproteinized U4/U6 purified from wildtype or *Δprp24sup* strains show that there both duplexes melt at ~54°C.

Glycerol gradient purified particle was deproteinized and incubated at the temperature listed. Both U4/U6 duplexes melt at ~54°C, as shown by the lack of U4/U6 band, and the appearance of a mono-U6 snRNA band. No significant difference between wildtype and *Δprp24sup* U4/U6 is apparent, suggesting that the absence of Prp24p does not result in major duplex destabilization.

### 4.3.6 Biochemical analysis of arrested di-snRNP complexes

U4/U6 is a transient assembly intermediate when compared with the assembly species that exist downstream and upstream of this snRNP. Because *in vivo* levels of U4/U6 are low, isolating this intermediate for compositional analysis has proven to be a difficult task. Previous studies have compositionally analyzed *in vitro* assembled U4/U6; however, these reconstitution assays do not necessarily represent authentic spliceosomal complexes, as they are not on-pathway intermediates. In addition, it is difficult to determine the timing of protein additions and removals from snRNP complexes when particles are assembled from extract.

Previously, a U4 snRNA cold-sensitive mutant, U4-G14C, had been characterized to stall the assembly pathway of U4/U6 at the Prp24p step (Shannon and Guthrie, 1991). This U4 mutant disrupts the base stacking of three G-C intermolecular U4/U6 base pairs, and leads to a destabilization between U4 and U6, which lowers the melting temperature of the duplex from 53°C to 37°C. Destabilizations such as this often lead to a temperature sensitive phenotype; however, this particular destabilization traps a complex containing U4, U6, and Prp24p in di-snRNP. In a wildtype background this intermediate containing U4, U6, and Prp24p is considered highly transient (Stevens *et al*., 2001). This led to the suggestion that this U4•Prp24p•U6 complex represents a functional intermediate along the assembly pathway (Shannon and Guthrie, 1991). Compositional characterization of the U4•Prp24p•U6 intermediate could provide insight a to whether Prp24p potentially acts in a di-snRNP complex in the presence of other di-snRNP factors or binds to di-snRNP so as to exclude other di-snRNP proteins from interacting prematurely.

#### *4.3.6.1 Affinity purification of Prp24p-containing di-snRNP*

To begin, Prp24p was C-terminally TAP tagged in a strain expressing a plasmid copy of U4-G14C and a genomic disruption of *SNR14*. Next, U4-Prp24p-TAP•U6 di-snRNP was affinity purified from cells which had been grown at permissive temperatures (31°C) and non-permissive temperatures (16°C). Following the affinity purifications (as described in Chapter 2, section 2.2.7), the elutions were layered on di-snRNP-resolving glycerol gradients and centrifuged to separate di-snRNP. Gradients were fractionated and RNA was extracted from protein. RNA was separated by denaturing electrophoresis and transferred to nylon membrane for Northern blotting. Northern blots were probed for all five snRNAs.

As can be seen in Figure 4.8A and B, U4 and U6 are both precipitated by Prp24p in this mutant at non-permissive and permissive temperatures, respectively, while a purification from a wildtype strain (Prp24p affinity tagged) precipitates only U6 snRNP (Stevens *et al*., 2001). The results also demonstrate that mono-U6 snRNA is present at a high sedimentation in the gradient consistent with previous findings that Prp24p immunoprecipitates U6 snRNP (Stevens *et al*., 2001). Additionally, no accumulation of snRNAs is apparent in the pellet fraction (fraction 28), indicating that tri-snRNP cannot be precipitated by Prp24p; a result consistent aligned with previous immunoprecipitations experiments (Stevens and Abelson, 1999; Stevens *et al*., 2001).

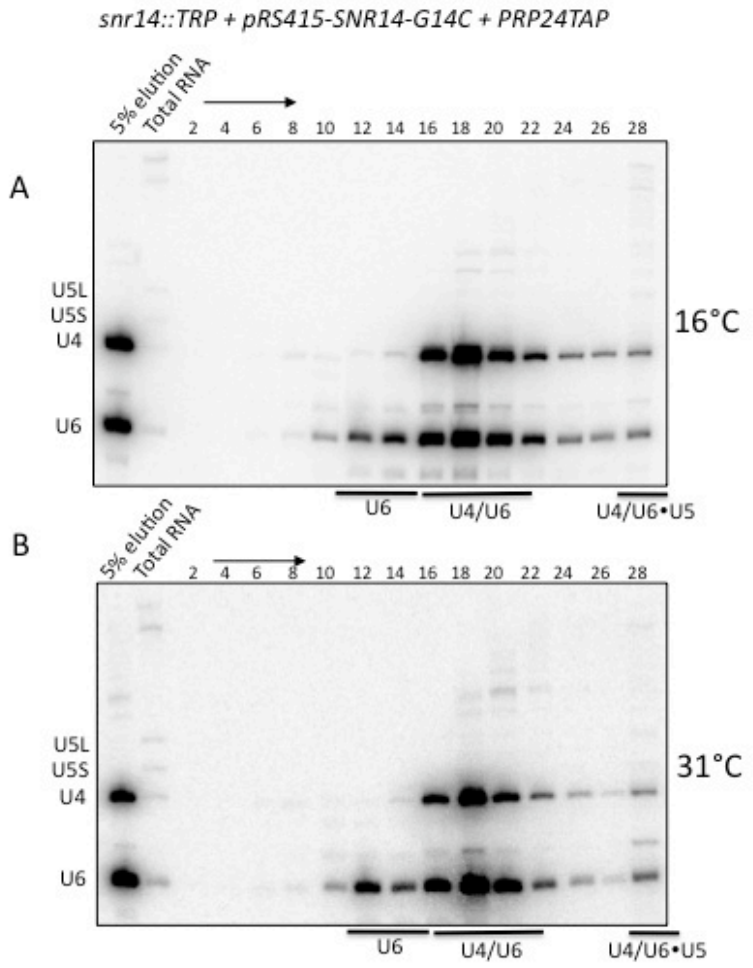


Figure 4.8 Prp24p immunoprecipitates both U4 and U6 in the when U4 is mutated.

At non-permissive (A) and permissive temperature (B) (16°C and 31°C, respectively), affinity purified U4 and U6 cosediment di-snRNP resolving glycerol gradients, indicating the presence of a di-snRNP particle containing Prp24p. This particle is suggested to be an intermediate in the assembly pathway of tri-snRNP (Jandrositz and Guthrie, 1995; Shannon and Guthrie, 1991).

#### *4.3.6.2 Stability of the interaction between U4 and U6 in di-snRNP*

The Guthrie laboratory reported that both U4 and U6 immunoprecipitated with Prp24p in the U4-G14C mutant strain (Shannon and Guthrie, 1991), posing the question of the mechanism of their interaction: is a protein, such as Prp24p, responsible for their interaction or do the two snRNAs interact by through base pairing interactions. Jandrositz and Guthrie use psoralen crosslinking and protein denaturation in nuclear extract to determine that U4 and U6 are in fact base paired in di-snRNP, indicating that their interaction is independent of proteins even in the presence of Prp24p (Jandrositz and Guthrie, 1995). Non-denaturing electrophoresis (Figure 4.9A) and psoralen crosslinking (Figure 4.9B) were used to study the extent of destabilization between U4 and U6.

Native RNA was extracted from strains containing either a wildtype copy of U4 or cells containing the U4-G14C mutant. As can be seen in lane 1 of Figure 4.9A, native extraction and non-denaturing electrophoresis of total RNA produces a band representing a U4/U6 base pairing interaction only in the strain carrying a wildtype copy of U4. Total RNA from the U4-G14C mutant shows no detectable amounts of U4/U6 base pairing (lane 3 of Figure 4.9A), suggesting a destabilization in the duplex that does not withstand standard native extraction and electrophoresis.

To determine if the lack of base pairing in the U4-G14C mutant strain is a result of the preparation of the RNA, a single glycerol gradient fraction of di-snRNP (from Figure 4.8 for the U4-C14 mutant, and Figure 4.5A for wildtype) was split and samples were incubated either in the presence of both UV and psoralen, in the presence of UV and absence of psoralen, in the absence of UV and presence of psoralen, or in the absence of both UV and psoralen. U4/U6 base pairing was expected only in the presence of both psoralen and UV. RNA from each of the samples was then extracted, precipitated and

subjected to electrophoresis on a denaturing gel. The gel was transferred to nylon, and the blot was probed for U6. As shown in Figure 4.9B, U4 and U6 do show a base pairing interaction only in the presence of both UV and psoralen, indicating that though the interaction is not able to withstand a standard native extraction, the two snRNAs are indeed participating in base pairing interactions. This is consistent with the data from the Guthrie laboratory, and indicates that even in the presence of Prp24p, U4 and U6 are base paired to some extent.

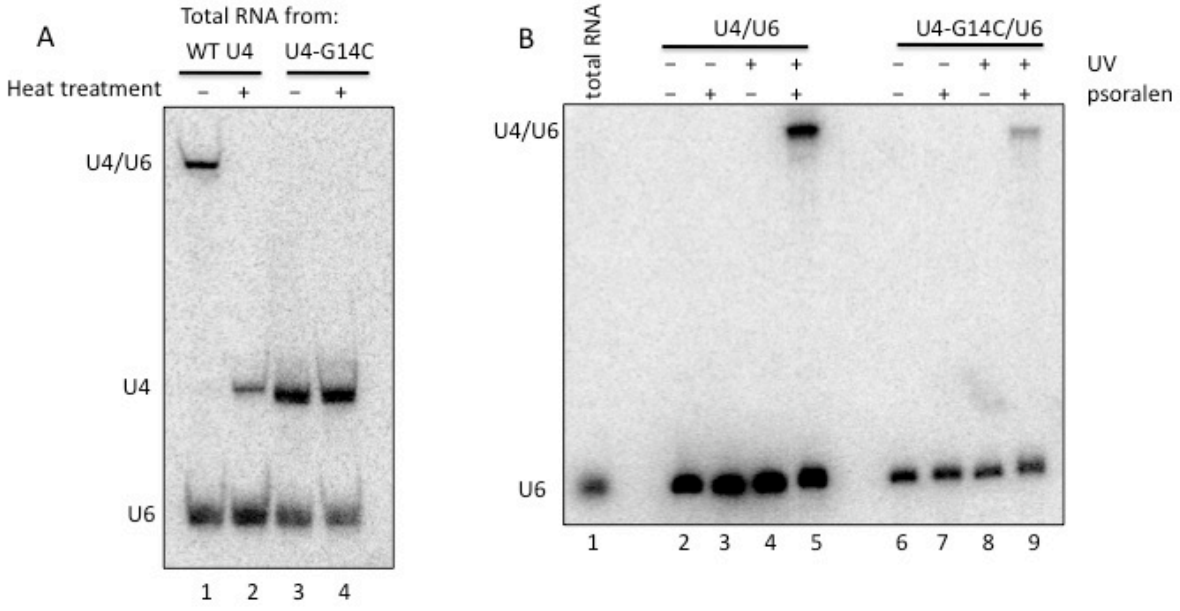


Figure 4.9 U4 and U6 are base paired in the U4-G14C•Prp24p•U6 particle

A. Non-denaturing electrophoresis and Northern blotting of total RNA from wildtype U4 and U4-G14C cells shows that U4 and U6 are stably base paired in wildtype U4 cells (lane 1), but not in U4-G14C cells (lane 3). RNA in lanes 2 and 4 was heat-treated as a control.

B. Glycerol gradient fractions were treated in the absence or presence of psoralen and UV light, and extracted RNA was separated on a denaturing gel. Northern blotting for U6 demonstrates that treatment with psoralen and UV light crosslinks U4 and U6 in affinity-purified di-snRNP from both wildtype (lane 5) and U4-G14C (lane 9) strains. In the absence of either psoralen or UV light, no base pairing is detected (lanes 2-4 and lanes 6-9). Lane 1 serves as an untreated total RNA control.

#### *4.3.6.3 Compositional analysis of Prp24p-containing di-snRNP*

Protein fractions corresponding to the RNA fractions where U4/U6 was detected (from section *4.3.6.1*) were precipitated, pooled and subjected to mass spectrometry. Table 4.5 shows that the only stable protein components present in this complex are Prp24p, the LSm core, and the Sm core. These results are not unexpected; however, it is interesting to note that no other di-snRNP components are present, indicating they may join in a mutually exclusive manner to Prp24p. It is also possible that the di-snRNP components cannot load into this complex when U4 and U6 are destabilized. These data suggest that the uncharacterized U4 snRNP, which enters into interactions with U6 snRNP does not carry or recruit other di-snRNP factors into di-snRNP, at least before it has formed an interaction with U6.

| Protein | MW (kDa) | Di-snRNP (# of peptides detected) |
|---|---|---|
| Prp24p | 51 | 24 |
| SmB1p | 22 | 15 |
| LSm4p | 21 | 11 |
| SmD1p | 16 | 11 |
| SmD2p | 13 | 9 |
| LSm7p | 13 | 3 |
| LSm3p | 12 | 11 |
| LSm8p | 12 | 15 |
| SmD3 | 11 | 8 |
| LSm2p | 11 | 9 |
| LSm5p | 10 | 12 |
| SmFp | 10 | 6 |
| SmEp | 10 | 1 |
| LSm6p | 9 | 9 |

Table 4.5 Factors present in Prp24p-purified di-snRNP include the Sm and LSm complex.

Compositional analysis of proteins found in Prp24p-containing di-snRNP was determined by mass spectrometry analysis. All of the core snRNP components of U4 and U6 snRNAs, the Sm and LSm complexes, respectively, are present. No other di-snRNP factors were identified suggesting they are recruited later in the assembly process, and are not brought in by the U4 snRNP.

#### *4.3.6.4 Compositional analysis of the di-snRNP contained in Δprp24sup strain*

Di-snRNP found in the *Δprp24sup* strain is base paired, as demonstrated in non-denaturing electrophoresis (Figure 4.5D), suggesting that this particle is likely to represent the post-Prp24p assembly intermediate. Additionally, as shown by native gel electrophoresis in Figure 4.4, di-snRNP migrates comparably different from di-snRNP from a wildtype native extract. Using the affinity-purified di-snRNP described in section 4.3.5.1, protein was precipitated, pooled and analyzed by mass spectrometry analysis to determine composition. Proteins identified are presented in Table 4.6. Indeed, this complex contains other di-snRNP factors, including Prp3p, Prp4p and Prp31p. Snu13p was not identified despite using an inclusion list specifically for Snu13p peptides; however, it cannot be ruled out that total protein is too low for detection of peptides from Snu13p. Evidence from the Lührmann laboratory suggests that this factor must join prior to Prp31p, Prp3p, and Prp4p forming interactions in the di-snRNP (Nottrott *et al*., 1999) (Nottrott *et al*., 2002).

Furthermore, as expected, this particle contains members of the Sm and LSm protein cores, as expected. Additionally, several identifications with significant scores were made of Brr2p, however, the large molecular weight of Brr2p makes resultant peptide identifications plausible even in the event that Brr2p does not specifically interact with the complex in question.

| Protein | Molecular weight (kDa) | *Δprp24sup* di-snRNP (# of peptides IDed) |
|---|---|---|
| Brr2p | 246 | 11 |
| Pab1p | 64 | 11 |
| Prp31p | 56 | 3 |
| Prp3p | 56 | 13 |
| Prp4p | 52 | 17 |
| LSm4p | 21 | 5 |
| SmD2p | 13 | 2 |
| LSm8p | 12 | 5 |
| LSm2p | 11 | 2 |
| LSm3p | 10 | 1 |
| SmFp | 10 | 1 |
| LSm6p | 9 | 1 |

Table 4.6 Compositional analysis of di-snRNP purified from *Δprp24sup* strain show that di-snRNP factors, Prp3p, Prp4p, and Prp31p join to the base paired duplex.

Mass spectrometry analysis of pooled protein from di-snRNP fractions of *Δprp24sup* purification identify di-snRNP factors not found in the Prp24p containing di-snRNP, suggesting that the Prp24p binds mutually exclusively to the other di-snRNP factors.

## 4.4 DISCUSSION

Previous studies have shown that Prp24p predominantly acts by structurally rearranging U6 snRNA into a conformation optimal for the formation of extensive base pairing interactions with the U4 snRNA. Recent studies have proposed a mechanism for this action whereby RRM2 recognizes precise residues in a bulged central stemloop of U6, thus bringing RRM1 of Prp24p into close enough proximity to capture U6 (Martin-Tumasz *et al*., 2010). It has not been determined if this rearrangement is concurrent or directly preceding U4 annealing, but Prp24p has not been shown to exist in complex with U4 snRNA in a wildtype background (Stevens *et al*., 2001).

Studies in yeast have shown that cells can compensate for a mutation in Prp24p by upregulating steady-state levels of U6 snRNA (Vidaver *et al*., 1999). This effect is conserved in vertebrates (Trede *et al*., 2007), underlining the significance of the observation. This finding begs the question of how increasing steady-state levels of U6 can compensate for defective Prp24p. While yeast and zebrafish knockouts of Prp24p are non-viable, we have demonstrated that by overexpressing U6 snRNA, viability can be restored in the absence of Prp24p. Interestingly, U4 and U6 can still base pair in the absence of Prp24p; however, Prp24p knockout strains display defects in the assembly of higher order spliceosomal intermediates. The results presented here suggest that the release of Prp24p may act as a trigger for the assembly of tri-snRNP.

Prp24p is well characterized to mediate the annealing between U4 and U6, but proposed roles for Prp24p function in later steps of the splicing cycle lack biochemical support. Using the Prp24p knockout strain, we have shown that the role of Prp24p extends beyond U4/U6 annealing. The data presented here suggest that Prp24p is

temporally important for tri-snRNP formation, and implying that the timing of Prp24p action is critical for cells to maintain a fully functional spliceosome biogenesis pathway.

We began by overexpressing either U6 snRNA or U4 snRNA in a *PRP24* shuffle strain to test the hypothesis that the essential nature of Prp24p can be bypassed by driving di-snRNP assembly process by mass action. Indeed, it was observed that elevated levels of U6 make Prp24p non-essential (Figure 4.1). Molecularly, the increased levels of U6 snRNA in this strain imply that there is a concomitant increase in the total amount of *de novo*-produced U6 snRNA entering the splicing cycle. As Prp24p has been defined as a recycling factor, it is possible that U6 snRNA produced *de novo* exists in a conformation that can readily anneal to U4 snRNA. Another possibility is that the increase in steady-state levels of U6 snRNA in a cell raises the levels of U6 snRNA in a conformation thermodynamically favorable for binding U4 snRNA. A combination of these effects is also possible.

While high levels of U6 snRNA can circumvent the requirement of Prp24p *in vivo*, non-denaturing electrophoresis demonstrated that cells lacking Prp24p could still form U4/U6 base pairing interactions, though the total amount of base paired duplex is reduced (Figure 4.2). The presence of base pairing is surprising considering the essential protein, Prp24p, has been assigned a functional role in annealing U4 and U6 (Raghunathan and Guthrie, 1998b). Thus, this result suggests that Prp24p may have an additional role *in vivo*. Instead of observing a complete absence of U4/U6 annealing, assembled di-snRNP and tri-snRNP in the *Δprp24sup* strain have unique migrations in a native gel when compared to di-snRNP and tri-snRNP assembled in wildtype cells (Figure 4.4). Taken together these data suggest that Prp24p may also be playing a direct or indirect role in facilitating downstream assembly events.

In the absence of Prp24p, a reduction in base pairing between U4 and U6 was observed; however, the interaction was not completely ablated. To confirm its role in the annealing of U4 and U6, Prp24p was metabolically depleted and the extent of base pairing between U4 and U6 was assessed. Following a shift to a repressive carbon source, Prp24p was not detectable by 4.5 hours (Figure 4.3B), and a concomitant decrease in the levels of U4 base paired to U6 was observed (Figure 4.3C). This *in vivo* data partially validates the previously reported function of Prp24p in U4/U6 annealing, but in light of our previous finding that Prp24p is not essential in the presence of high levels of U6, our results imply that Prp24p is not essential for the interaction. This scenario provides an appealing explanation for the finding that high levels of U6 can bypass the need for Prp24p.

To test whether the migration differences noted in native gel electrophoresis of the *Δprp24sup* spliceosomal particles (Figure 4.4) are the result of a previously undescribed function for Prp24p in later spliceosome assembly events, we affinity purified the tri-snRNP and di-snRNP assembled *in vivo* in the *Δprp24sup* strain. Di-snRNP and tri-snRNP were purified by means of a C-terminal TAP tag on LSm8p in the *Δprp24sup* strain, and large-scale affinity purifications were performed. A comparison of the protein composition of tri-snRNPs from wildtype and *Δprp24sup* revealed very few compositional differences. The presence of Cbf5p in the *Δprp24sup* tri-snRNP is notable, as Cbf5p is a component of the box H/ACA snoRNA, which pseudouridylates U2 snRNA during U2 snRNP biogenesis (Ma *et al*., 2005). Other snRNAs have been reported to be modified by pseudouridylation, as well (Massenet *et al*., 1998); however, Cbf5p has only been specifically reported for modifying U2.

More interesting is the absence of Prp38p and Snu23p in the *Δprp24sup* tri-snRNP. Both of these factors are reported to be stable components of tri-snRNP (Stevens

and Abelson, 1999), and their absence is surprising. Prp38p has been reported to be important for unwinding of U4 snRNA during activation (Xie *et al*., 1998). It is interesting to note that the presence of Prp38p, a factor important for unwinding, appears to be dependent upon the presence of another factor, namely Prp24p, proposed to function in the same activation event. However, the purifications of tri-snRNP described here do not answer whether Prp24p directly affects the stability of Prp38p.

Since no significant compositional changes were observed between wildtype tri-snRNP and tri-snRNP assembled in *Δprp24sup* strain, the migration shift in the native gel is likely to be due to conformational changes in the U4/U6 interaction. One approach to test this hypothesis is to study the thermal stability of the U4/U6 duplex in the presence and absence of Prp24p. Melting experiments revealed that the U4/U6 interaction in *Δprp24sup* tri-snRNP is similar to that of wildtype tri-snRNP, suggesting that reduced migration in the native gel is not due to incomplete or destabilized base pairing interactions between U4 and U6. Native gel analysis separates by both size and charge, thus it is possible that the tri-snRNP produced from *Δprp24sup* strains is structurally different or that there are fewer exposed charged residues than that from wildtype cells. In either case, these data suggest that Prp24p is important for either tri-snRNP assembly or stability, or both, and implicate a previously uncharacterized function for Prp24p in later spliceosome assembly events.

One of the possibilities for the aberrantly migrating tri-snRNP in the *Δprp24sup* strain is that Prp24p may be functioning in assembly of the spliceosome indirectly. The role of Prp24p in di-snRNP formation may be kinetically or temporally significant, as other factors are known to eventually associate with di-snRNP on the pathway to tri-snRNP. To investigate the effect that knocking out Prp24p has on di-snRNP, proteins specific to di-snRNP were identified by mass spectrometry analysis, and compared to di-

snRNP isolated from cells containing a mutation U4 snRNA that has been shown to accumulate a Prp24p-containing di-snRNP (Shannon and Guthrie, 1991). In the presence of Prp24p, di-snRNP is composed of only Sm and LSm complexes, while di-snRNP purified from *Δprp24sup* cells contains other di-snRNP components, including Prp3p, Prp4p, and Prp31p, in addition to the Sm and LSm complexes. The absence of Snu13p may be significant; however, Snu13p is not easily identified using this method of mass spectrometry analysis even in the presence of a Snu13p-specific inclusion list. Additionally, Snu13p is thought to recruit the other factors to di-snRNP (Nottrott *et al*., 1999; Oruganti *et al*., 2005), thus the absence of Snu13p is likely an artifact of the method of identification.

The two di-snRNP characterizations suggest that Prp24p may dissociate from di-snRNP upon the joining of other factors. Perhaps these factors physically remove Prp24p from di-snRNP, or it is possible that the other factors induce rearrangements in the snRNAs that cause the Prp24p interaction to be disrupted. In either case, it appears that Prp24p precedes the entrance of the other di-snRNP factors into the assembly pathway, as opposed to their concomitant recruitment as possible U4-interacting factors.

Finally, as noted above, Prp24p has been implicated in dissociation of U4 during activation of the spliceosome (Ghetti *et al*., 1995; Vidaver *et al*., 1999), but the viability of the *Δprp24sup* strain containing high levels of U6 snRNA indicate that if Prp24p does in fact play a role in activation, it is not an essential function for Prp24p, or less likely that it too can be bypassed with high levels of U6 snRNA. Prp24p is essential for the recycling of U6 snRNA, but the data proves that this function for Prp24p can be bypassed by increasing the cellular levels of U6. Though this strain has a slight growth defect, the fact that cells are viable proves that all of functions of Prp24p are non-essential, at least in the presence of high copy U6.

# Chapter 5: Significance and future directions

## 5.1 SIGNIFICANCE

As a highly dynamic RNP, the spliceosome presents a uniquely challenging system in understanding the network of RNA and protein interactions. Biochemical purifications of long-lived spliceosomal intermediates have characterized novel factors and interactions, but because the nature of the splicing cycle is largely in flux, *in vitro* strategies for characterizing the more transient-existing intermediates have remained the major source of data. Though these *in vitro* techniques are invaluable to mechanistic studies, the particles produced from *in vitro* reconstitutions are not assembled along the splicing cycle pathway, and thus are not necessarily the most accurate method for understanding a dynamic RNP. Additionally, the particles produced are not always functional due to the method of isolation and the fact that large populations of reconstituted particles are off-pathway, dead-end intermediates. Thus, genetic strategies for arresting normally short-lived species of splicing particles, *in vivo,* provide an essential means by which the spliceosome must be studied.

We have employed these genetic strategies to accumulate normally transient spliceosomes, and solve the compositions of each of these particles. Furthermore, we have shown that given the appropriate conditions these arrested particles are also functional. By testing the functionality of the catalytically arrested spliceosomes, we have assigned a mechanistic function to a DEAH-box factor, Prp2p, and found a possible secondary consequence of Prp2p action. Additionally, using genetic strategies, we have shown that assembly intermediates can be accumulated and analyzed by using assembly-factor mutants. By linking genetic and biochemical experiments, the data presented in

this dissertation provides a means to compositional characterizations of the assembly intermediates. From the results, we can conclude that the factor, Prp24p, previously narrowed to U6 rearranging, also participates in spliceosomal assembly events downstream of U4/U6 annealing. Moreover, we show that Prp24p is not essential in activation of the spliceosome, an event previously ascribed to Prp24p.

Combining these genetic principles and biochemical methods we present a means by which snapshots of the dynamic splicing cycle can be mechanistically pieced together. Nevertheless there is still much that remains to be learned about how the spliceosome functions in molecular detail. In the following, possible future directions based on the work described in this dissertation will be discussed.

## 5.2 CATALYTICALLY ARRESTED SPLICEOSOMES

The second chapter of this dissertation describes the genetic approach used for isolating normally transiently occurring intermediates in the splicing cycle. Specifically, catalytically activated spliceosomes arrested before the first step, before the second step, and before message released were purified and compositionally defined by mass spectrometry analysis. These analyses revealed a number of interesting protein-addition and protein-removal events. The third chapter elaborates upon one of these compositional changes using functional analysis to show a mechanistic role for Prp2p in the first step of splicing. Further analyses of the compositional differences will be instrumental in assigning functions to uncharacterized proteins present during the catalytic steps of splicing.

### 5.2.1 Compositional changes between arrested spliceosomes

Our studies demonstrate that there are many noticeable differences between each of the arrested spliceosomes, suggesting that the conformational rearrangements and chemical steps of splicing trigger protein additions and removals from the spliceosome. We specifically investigated one of these removals. We have tested the functionality of the spliceosome to carry out the first step of splicing, and we found that the SF3 complex is quantitatively displaced from the spliceosome as a result of ATP-hydrolysis by Prp2p. In addition to the displacement of the SF3 complex, we also identified numerous other protein and complex changes, suggesting that they too might be displaced or added following ATPase function. One of the other removals from the *prp2*-arrested spliceosome is the RES complex, which is known to associate with the pre-mRNA to prevent the message from entering into the cytoplasm prematurely (Dziembowski *et al*., 2004). However, the mechanism that triggers the removal of this complex has not been established. A strategy similar to that described for the displacement of the SF3 proteins may allow determination of whether the RES complex or any of the other proteins specific to *prp2*-arrested spliceosomes are removed as a direct consequence of the DEAH-box proteins.

Now that at least one consequence of Prp2p function has been established, it will be interesting to determine how Prp2p elicits displacement of the SF3 complex. Previous results strongly support Prp2p recruitment to the spliceosome by Spp2p (Roy *et al*., 1995); however, where Spp2p and Prp2p interact within the spliceosome remains unknown. Our arrested particles provide a system by which protein-RNA interactions can be investigated by Crosslinking Analysis of cDNA (CRAC) (Granneman *et al*., 2009). This recently described technique takes advantage of the specific RNA-protein crosslinks produced by UV light. Specifically, an affinity-purified particle is exposed to UV light,

and the particle is treated with RNases. A denaturing purification of the digested RNP, specific to the protein interest is then employed. While bound on resin, linkers are then ligated to the RNA of interest on both the 5' and 3' ends, which can eventually be used for primer annealing sites for reverse transcription and then PCR. This amplifies the RNA(s) bound by the protein of interest, and can be used in a high-throughput manner. A secondary tag within the mutant *prp2* strain, on either Prp2p or potentially RNA-binding protein of interest, such as Spp2p, may be used for a second purification step of the digested RNP to identify the exact points of interaction, whether the interaction is with an snRNA or the pre-mRNA.

Using the CRAC strategy on our purified arrested spliceosomes would provide a snapshot of interactions among the factors that remain part of the spliceosome throughout the catalytic steps. Proteins of interest within the first and second step-arrested particle include Prp8p, Brr2p, and Snu114p. Prp8p, in particular, is considered by at the heart of the active site of the spliceosome, thus, understanding the interactions between the snRNAs, the pre-mRNA, and Prp8p within each of the catalytically arrested particles would help elucidate the mechanistic role of Prp8p in the steps of splicing.

Comparison between the compositions of the first-, second-, and message release-arrested spliceosomes also reveal a number of protein additions. For example, Yju2p and Cwc25p are only found in the *prp16*-arrested spliceosome suggesting that they are added to the spliceosome sometime after Prp2p function. Our functional analysis suggests that neither Yju2p nor Cwc25p are essential for the first step of catalysis; however, both of these proteins have been demonstrated to be important for the first step using *in vitro*-assembled first step-stalled spliceosomes (Chiu *et al*., 2009; Liu *et al*., 2007; Warkocki *et al*., 2009). Determination of where Yju2p and Cwc25p bind in the second step arrested spliceosome by CRAC may help elucidate the purposes of these factors. Specifically, the

primary tag on *prp16* and a secondary tag on either Cwc25p or Yju2p could reveal the location of these factors within the spliceosome. We predict that these factors may be interacting on and/or around the second step reactive species before the second step of catalysis.

### 5.2.2 Analysis of other particles

Using genetic strategies to isolate normally transient occurring intermediates, provides the means to isolate and analyze *in vivo* assemble, synchronized, on-pathway, functional intermediates. As described in the Chapter 4, genetic strategies can be used to isolate assembly intermediates as well. This strategy is valuable in addressing the specific consequences of early assembly intermediates affecting later spliceosome assembly intermediates and would be useful if applied to other particles, such A complex spliceosomes or activated B* spliceosomes, *in vivo*.

## 5.3 CHARACTERIZING THE ASSEMBLY PATHWAY

Prp24p has been well characterized as a mono-U6 snRNP component (Shannon and Guthrie, 1991; Stevens *et al*., 2001), and the traditional role of Prp24p is to anneal U4 and U6 in one of the earliest assembly pathways (Raghunathan and Guthrie, 1998b). This event may take place by simply relaxing the U6 secondary structure such that U4 has the opportunity to form base pairing interactions with U6. It is also possible that Prp24p, which contains four RRMs, actually binds U4 snRNA in addition to U6, and that this U4-Prp24p interaction has never been isolated due to the transient nature of this particle. Using our *in vivo*-assembled, Prp24p containing di-snRNP, an interaction for U4-Prp24p can be explored using the CRAC method described above. An interaction

between U4 and Prp24p would help elucidate the formation of the di-snRNP and the mechanism by which Prp24p anneals U4 and U6 snRNAs.

Though our data clearly show that Prp24p is not required for the activation of the spliceosome, Prp24p may be involved in stimulating the unwinding of U4 snRNA in a non-essential manner. It would be interesting to further investigate this potential role for Prp24p by biochemically exploring the existence of Prp24p in a spliceosome-sized particle. Previous compositional studies have shown the exclusive binding of Prp24p to U6 snRNA (Stevens *et al*., 2001); however, the transient nature of some of the assembly intermediates may not permit biochemical isolation. Using the U4-G14C strain, which is known to be defective for Prp24p release, a spliceosome component, such as Brr2p, could be affinity tagged. An affinity purification following glycerol gradient sedimentation specific for a spliceosome-sized particle, may reveal that Prp24p is prevented from release not only during tri-snRNP formation, but also during the activation event, implying a function for Prp24p in a fully assembled, B-complex spliceosome. Though several reports have provided genetic evidence of a role for Prp24p in the activation step, *in vivo* biochemical data would provide physical evidence for a non-essential role for Prp24p just upstream of the catalytic steps of splicing.

Finally, mass spectrometry analysis of tri-snRNP produced from the *Δprp24sup* strain, identified the snoRNP component Cbf5p. Cbf5p is known to be involved in pseudouridylation of U2 snRNA (Ma *et al*., 2005), and although tri-snRNP by definition does not contain the U2 snRNP, it is possible that Cbf5p is involved in modifying one or all of the other snRNAs. Using a TAP tag on Cbf5p in the *Δprp24sup* strain, it would be interesting to see if spliceosomal intermediates co-precipitate.

# Appendix I

Mass spectrometry protocol used in Chapter 2 to compositionally characterize arrested spliceosomes as described by James Thompson of the Yates Laboratory

To each lyophilized protein sample, 60 µL of solubilization buffer (8M urea, 10 mM Tri-HCl, pH 8.5) was added. The subsequent mixture was then reduced by adding TCEP to 5 mM. Iodoacetamide (10 mM final concentration) was added to alkylate cysteine residues and the samples were subsequently incubated at room temperature in the dark for 15 minutes. Reactions were diluted to 2 M urea by the addition of 180 µL of 100 mM Tris, pH 8.5. Calcium chloride was added to a final concentration of 1 mM. Two micrograms of Trypsin were added and the resulting mixtures were shaken for 18 hours in the dark at 37°C. Formic acid (90%) was added to a final concentration of 5% formic acid. The tubes were centrifuged for 30 minutes at 2°C on a table-top centrifuge.

The protein digest was pressure-loaded onto a fused silica capillary desalting column containing 5 cm of 5 µm Polaris C18-A material (Metachem) packed into a 250 µm i.d. capillary with a 2 µm filtered union (UpChurch Scientific). The desalting column was washed with buffer containing 95% water, 5% acetonitrile, and 0.1% formic acid. After desalting, a 100 µm i.d. capillary with a 5µm pulled tip packed with 10 cm 3-µm Aqua C18 material (Phenomenex) followed by 3 cm of 5 µm Partisphere strong cation exchanger (Whatman) was attached to the filter union and the entire split-column (desalting column-filter union-analytical column) was placed inline with an Agilent 1100 quaternary HPLC and analyzed using a modified six-step separation described previously (Washburn *et al*., 2001). The buffer solutions used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B), and 500 mM ammonium acetate/5%acetonitrile/0.1% formic acid (buffer C). Step 1 consisted of a 90 minute

gradient from 0 – 100% buffer B. Steps 2-5 had the following profile: 3 minutes of 100% buffer A, 2 minute X% buffer C a 10 minutes gradient from 0-15% buffer B and a 97 minute gradient from 15-45% buffer B. The 2 minute buffer C percentages (X) were 20%, 40%, 60%, and 80%, respectively, for the seix step analysis. For the final step, the elution consisted of: 3 minutes of 100% buffer A, 20 minutes of 100% buffer C, a 10 minute gradient from 0-15% buffer B and a 107 minute gradient from 15-70% buffer B.

As peptides were eluted from the microcapillary column, they were electrosprayed directly into an LTQ two-dimensional ion trap mass spectrometer (ThermoFinnigan) with the application of a distal 2.4-kV spray voltage. A cycle of one full-scan mass spectrum (400-1400 m/z) followed by eight data-dependent MS/MS spectra at a 35% normalized collision energy was repeated continuously throughout each step of the multidimensional separation. Application of mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system.

MS/MS spectra were analyzed using the following software analysis protocol. Poor quality spectra were removed from the data set using an automated spectral quality assessment algorithm (Bern *et al*., 2004). MS/MS spectra remaining after filtering were searched with the SEQUEST algorithm (Eng *et al*., 1994) against the SGD *S. cerevisiae* (December 16, 2005) protein database (created on December 17, 2006) concatenated to a decoy database in which the sequence for each entry in the original database was performed on a Beowulf computer cluster consisting of 100 1.2 GHz Athlon CPUs (Sadygov *et al*., 2002). No enzyme specificity was considered for any search. SEQUEST results were assembled and filtered suing the DTASelect (version 2.0) program (Cociorva *et al*., 2006; Tabb *et al*., 2002). DTASelect 2.0 uses a linear discriminate analysis to dynamically set XCorr and DeltaCH thresholds for the entire data set to achieve a user-

specified false-positive rate. The false-positive rates are estimated by the program from the number and quality of spectral matches to the decoy database.

# Appendix II

Contaminants from mass spectrometry analysis of mock purification in Chapter 2.

| **MOCK control** | | |
|---|---|---|
| Cdc19p | Hor2p | Shm2p |
| Pgk1p | Hyr1p | Egd2p |
| Trx2p | Sse1p | Rnq1p |
| Fba1p | Cdc48p | Frm2p |
| Tef2p, Tef1p | Cof1p | Glk1p |
| Tsa1p | Ssc1p | Gpd1p |
| Eno2p | YDL124W | Efg20p |
| Eno1p | Mmf1p | Adk1p |
| Ahp1p | YBR012W-A | Arg1p |
| Gpm1p | YDR098C-A | Gcv3p |
| Mbf1p | YDR210C-C | Leu1p |
| Hsp26p | YDR261C-C | Wtm1p |
| Pdc1p | YMR051C, YDR316W | Aro4p |
| Bmh1p | Gre2p | Aro8p |
| Hmf1p | Sam2p | Sis1p |
| Tal1 | Dps1p | Lys21p |
| Tpi1p | Eft1p, Eft2p | Pfk2p |
| Tma19p | Hyp2p | Scl1p |
| Pgi1p | Act1p | Stm1p |
| Grx5p | YMR090W | Cpr6p |
| Trx1p | Hsp104p | Hxk2p |
| Hhf2p, Hhf1p | Hhp31p | Ipp1p |
| Rhr2p | Cpr3p | Erg13p |
| Sod1p | Imd4p | Ser3p |
| Ynk1p | Tif2p, Tif1p | Ser33p |
| Bmh2p | YLR179C | Aha1p |
| Ado1p | Gly1p | Ald6p |
| Ntf2p | Oye2p | Gnd1p |
| Hsc82p | Guk1p | Hem2p |
| Trr1p | Pre9p | Tfp1p |
| YNL134C | Cys4p | Hom6p |
| Imd3p | Tkl1p | Tef4p |
| Yef3 | Cpr1p | Fas2p |
| Pnc1p | Hsp60p | Hsp82p |
| Sah1p | Nap1p | |

# Appendix III

Contaminating polypeptides from arrested particles characterized in Chapter 2.

| ***prp2*-arrested** | ***prp16*-arrested** | ***prp22*-arrested** |
|---|---|---|
| Atp1p | Air1p | Apa1p |
| Atp2p | Arp9p | Arf1p |
| Bgl2p | Atp1p | Arp7p |
| Csn12p | Cdc33p | Atp2p |
| Kri1p | Eaf3 | Cdc22p |
| Krr1p | Ebp2p | Dbp6p |
| Lat1p | Hsp10p | Drs1p |
| Loc1p | Ilv3p | Hpr1p |
| Mpd2p | Kre33p | Krr1p |
| Nop12p | Med1p | Mft1p |
| Nop13p | Med6p | Nog1p |
| Pma2p | Nog1p | Nop1p |
| Rpf2p | Nop1p | Nop58p |
| Sbp1p | Nop58p | Por1p |
| | Rfc1p | Sbp1p |
| | Rpc19p | |
| | Rsc4p | |
| | Swi1p | |

# Appendix IV

Mass spectrometry analysis of assembly intermediates characterized in Chapter 4,

Protein digests were preformed by Michelle Gadush at the Institute for Cellular and Molecular Biology Protein Microanalysis Facility. Protein pellets were dissolved in 25 μL ultrapure, filtered water. To the dissolved pellet, 5 μL of denature solution was added (8 M urea, 100 μM DTT, 60 μM $CaCl_2$, 64 mM $NH_4HCO_3$). The solutions were allowed to denature at room temperature for no less than one hour. To digest the samples, 10 μL of trypsin solution (Promega) was added (100 ng total enzyme), and digests were left shaking at 37°C overnight. To the digests, 3 μL of 5 mM DTT and 7 μL of 10% TFA were added.

Samples number corresponds to the following particles:

2. Compositional analysis of wildtype Prp3p-TAP purified tri-snRNP

4. Compositional analysis of tri-snRNP from strain lacking Prp24p

5. Compositional analysis Prp24p-containing di-snRNP

6. Compositional analysis of di-snRNP from strain lacking Prp24p

Protein identification was performed in the ICMB Proteomics and Metabolomics Facility by Dr. Maria D. Person (Dennis *et al*., 2009; Gorini *et al*., 2010). Tryptic peptides were loaded onto a Dionex Switchos μ-precolumn trap cartridge (Acclaim Peptrap C18 precolumn) flowing buffer A at 30 μl/min and then separated on an Dionex Ultimate HPLC flowing at 240 nL/min with a 73 (sample 2), 80 (4) or 100 (5,6) min gradient from 5-60% B (2) or 5-65% B (4-6). Buffer A consists of 0.05% TFA, 5%

acetonitrile, 95% water and Buffer B is 0.04% TFA, 95% acetonitrile, 5% water. The HPLC column used was the Dionex Acclaim PepMap 100 C18 column, with 3 μm, 100 Å beads, 75 μm inner diameter x 15 cm. The samples were spotted with a Dionex Probot spotting robot using 30 sec per spot, on 96 spots (2) or the whole plate of a 192 well target (4-6). α-cyano-4-hydroxycinnamic acid matrix (LaserBio Labs) made at 10 mg/ml in 80% acetonitrile was deposited simultaneously at a rate of 1.08 μl/min. Target spotting began 12-23 minutes after the start of the HPLC run.

MS and MS/MS were acquired on a MALDI-TOF/TOF (AB Sciex 4700 Proteomics Analyzer) running 4000 Series Explorer V3.6. MS was acquired over a mass range of 800-4000. The LC interpretation method acquired up to 12 MS/MS per spot from peaks with MS s/n > 30. Samples 3-6 had MS/MS exclusion lists with 35 keratin peaks and 65 PRP8 peaks (4). An MS/MS inclusion list was used including peptides from proteins Snu13 and 9 (4) or 14 (5,6) other small Sm and LSm proteins (LSm 2,3,4,5,7,8; Snu23, SmE, SmF, Dib1, SP381, PRP38, SmD2 and SmB). The spectra were processed using GPS Explorer V3.6 with MASCOT V2.2 as the embedded search engine.

The data was searched against the Swiss Prot (January 19, 2010) all species database and a yeast only subset database. The search parameters included trypsin/P with up to 2 missed cleavages, MS mass tolerance was set at 40 or 50 ppm, MS/MS mass tolerance of 0.2 Da, and the following variable modifications were specified: protein N-terminal acetylation and methionine oxidation. The MASCOT scores were reported according to standard scoring. The search output combines the scores from MS search and the MS/MS search using a probabilistic MOWSE algorithm. The MASCOT score is

defined as –10*logP, where P is the probability that the observed match is a random event. A score which corresponds to $P<0.05$ is chosen as the cutoff for a significant hit. For the Swiss Prot all species search, a score of greater than 40 for individual MS/MS was the significant cutoff. Human keratin proteins and bovine serum albumin carryover were detected in the all species search, along with yeast complex proteins. The yeast database search was used to restrict the identifications to complex proteins.

# Appendix V

Protein contaminants from mass spectrometry analysis of assembly intermediates characterized in Chapter 4.

| **Wildtype tri-snRNP** | **Tri-snRNP from *prp24sup*** | **Prp24p (U4G14C) di-snRNP** | **Di-snRNP from *prp24sup*** |
|---|---|---|---|
| Hsp75p | Rnq1p | Hsp75p | Rtg2p |
| | Rps5p | Hsp72p | Fun12p |
| | Pab1p | Adh1p | Hsp71p |
| | | Tdh3p | Tdh3p |

# References


Abruzzi KC, Lacadie S, Rosbash M (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J* **23:** 2620-31.

Abu Dayyeh BK, Quan TK, Castro M, Ruby SW (2002). Probing interactions between the U2 small nuclear ribonucleoprotein and the DEAD-box protein, Prp5. *J Biol Chem* **277:** 20221-33.

Achsel T, Ahrens K, Brahms H, Teigelkamp S, Luhrmann R (1998). The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol* **18:** 6756-66.

Achsel T, Brahms H, Kastner B, Bachi A, Wilm M, Luhrmann R (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* **18:** 5789-802.

Anthony JG, Weidenhammer EM, Woolford JL, Jr. (1997). The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA* **3:** 1143-52.

Arenas JE, Abelson JN (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A* **94:** 11798-802.

Ares M, Jr., Grate L, Pauling MH (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5:** 1138-9.

Ayadi L, Callebaut I, Saguez C, Villa T, Mornon JP, Banroques J (1998). Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J Mol Biol* **284:** 673-87.

Azubel M, Habib N, Sperling R, Sperling J (2006). Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. *J Mol Biol* **356:** 955-66.

Azubel M, Wolf SG, Sperling J, Sperling R (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol Cell* **15:** 833-9.

Bae E, Reiter NJ, Bingman CA, Kwan SS, Lee D, Phillips GN, Jr. *et al* (2007). Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J Mol Biol* **367:** 1447-58.

Banroques J, Abelson JN (1989). PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol* **9:** 3710-9.

Behrens SE, Tyc K, Kastner B, Reichelt J, Luhrmann R (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol* **13:** 307-19.

Bennett M, Michaud S, Kingston J, Reed R (1992). Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev* **6:** 1986-2000.

Berget SM, Moore C, Sharp PA (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74:** 3171-5.

Berglund JA, Fleming ML, Rosbash M (1998). The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* **4:** 998-1006.

Bern M, Goldberg D, McDonald WH, Yates JR, 3rd (2004). Automatic quality assessment of peptide tandem mass spectra. *Bioinformatics* **20 Suppl 1:** i49-54.

Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452:** 846-50.

Bjorn SP, Soltyk A, Beggs JD, Friesen JD (1989). PRP4 (RNA4) from Saccharomyces cerevisiae: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol* **9:** 3698-709.

Blanton S, Srinivasan A, Rymond BC (1992). PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. *Mol Cell Biol* **12:** 3939-47.

Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervey D (2009). Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* **36:** 583-92.

Bordonne R (2000). Functional characterization of nuclear localization signals in yeast Sm proteins. *Mol Cell Biol* **20:** 7943-54.

Bordonne R, Banroques J, Abelson J, Guthrie C (1990). Domains of yeast U4 spliceosomal RNA required for PRP4 protein binding, snRNP-snRNP interactions, and pre-mRNA splicing in vivo. *Genes Dev* **4:** 1185-96.

Botstein D, Fink GR (1988). Yeast: an experimental organism for modern biology. *Science* **240:** 1439-43.

Boudvillain M, de Lencastre A, Pyle AM (2000). A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature* **406:** 315-8.

Brody E, Abelson J (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228:** 963-7.

Brosi R, Groning K, Behrens SE, Luhrmann R, Kramer A (1993a). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262:** 102-5.

Brosi R, Hauri HP, Kramer A (1993b). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* **268:** 17640-6.

Brow DA, Guthrie C (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* **334:** 213-8.

Burgess S, Couto JR, Guthrie C (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* **60:** 705-17.

Burgess SM, Guthrie C (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73:** 1377-91.

Busch H, Reddy R, Rothblum L, Choi YC (1982). SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem* **51:** 617-54.

Caspary F, Seraphin B (1998). The yeast U2A'/U2B complex is required for pre-spliceosome formation. *EMBO J* **17:** 6348-58.

Caspary F, Shevchenko A, Wilm M, Seraphin B (1999). Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. *EMBO J* **18:** 3463-74.

Chan SP, Cheng SC (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem* **280:** 31190-9.

Chan SP, Kao DI, Tsai WY, Cheng SC (2003). The Prp19p-associated complex in spliceosome activation. *Science* **302:** 279-82.

Chapman KB, Boeke JD (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* **65:** 483-92.

Chen CH, Yu WC, Tsao TY, Wang LY, Chen HR, Lin JY *et al* (2002). Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Res* **30:** 1029-37.

Chen JH, Lin RJ (1990). The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. *Nucleic Acids Res* **18:** 6447.

Chen JY, Stands L, Staley JP, Jackups RR, Jr., Latus LJ, Chang TH (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell* **7:** 227-32.

Chen Y, Potratz JP, Tijerina P, Del Campo M, Lambowitz AM, Russell R (2008). DEAD-box proteins can completely separate an RNA duplex using a single ATP. *Proc Natl Acad Sci U S A* **105:** 20203-8.

Chen YI, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW (2007). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res* **35:** 3928-44.

Chiu YF, Liu YC, Chiang TW, Yeh TC, Tseng CK, Wu NY *et al* (2009). Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol Cell Biol* **29:** 5671-8.

Cho EJ, Takagi T, Moore CR, Buratowski S (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11:** 3319-26.

Chow LT, Gelinas RE, Broker TR, Roberts RJ (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12:** 1-8.

Church GM, Gilbert W (1984). Genomic sequencing. *Proc Natl Acad Sci U S A* **81:** 1991-5.

Cociorva D, Tabb DL, III YJ (2006). Validation of tandem mass spectrometry database search results using DTASelect. *Curr Protoc Bioinformatics* **16:** 13.4.1-13.4.14.

Colot HV, Stutz F, Rosbash M (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev* **10:** 1699-708.

Combs DJ, Nagel RJ, Ares M, Jr., Stevens SW (2006). Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis. *Mol Cell Biol* **26:** 523-34.

Company M, Arenas J, Abelson J (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349:** 487-93.

Cooper TA, Wan L, Dreyfuss G (2009). RNA and disease. *Cell* **136:** 777-93.

Cordin O, Tanner NK, Doere M, Linder P, Banroques J (2004). The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* **23:** 2478-87.

Davis CA, Grate L, Spingola M, Ares M, Jr. (2000). Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* **28:** 1700-6.

de la Cruz J, Kressler D, Linder P (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem Sci* **24:** 192-8.

Dennis MD, Person MD, Browning KS (2009). Phosphorylation of plant translation initiation factors by CK2 enhances the in vitro interaction of multifactor complex components. *J Biol Chem* **284:** 20615-28.

Domdey H, Apostol B, Lin RJ, Newman A, Brody E, Abelson J (1984). Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* **39:** 611-21.

Dougherty WG, Cary SM, Parks TD (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171:** 356-64.

Dunn EA, Rader SD (2010). Secondary structure of U6 small nuclear RNA: implications for spliceosome assembly. *Biochem Soc Trans* **38:** 1099-104.

Dybkov O, Will CL, Deckert J, Behzadnia N, Hartmuth K, Luhrmann R (2006). U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol Cell Biol* **26:** 2803-16.

Dziembowski A, Ventura AP, Rutz B, Caspary F, Faux C, Halgand F *et al* (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* **23:** 4847-56.

Edwalds-Gilbert G, Kim DH, Kim SH, Tseng YH, Yu Y, Lin RJ (2000). Dominant negative mutants of the yeast splicing factor Prp2 map to a putative cleft region in the helicase domain of DExD/H-box proteins. *RNA* **6:** 1106-19.

Edwalds-Gilbert G, Kim DH, Silverman E, Lin RJ (2004). Definition of a spliceosome interaction domain in yeast Prp2 ATPase. *RNA* **10:** 210-20.

Eng J, McCormack A, Yates JR, 3rd (1994). An approach to correlate tandem mass psectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5:** 976-989.

Enger MD, Walters RA (1970). Isolation of low molecular weight, methylated ribonucleic acids from 10S to 30S particles of Chinese hamster cell fractions. *Biochemistry* **9:** 3551-62.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H *et al* (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36:** 593-608.

Fleckner J, Zhang M, Valcarcel J, Green MR (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* **11:** 1864-72.

Fong N, Bentley DL (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15:** 1783-95.

Fortes P, Kufel J, Fornerod M, Polycarpou-Schwarz M, Lafontaine D, Tollervey D *et al* (1999). Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Mol Cell Biol* **19:** 6543-53.

Fortner DM, Troy RG, Brow DA (1994). A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev* **8:** 221-33.

Ganot P, Jady BE, Bortolin ML, Darzacq X, Kiss T (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* **19:** 6906-17.

Gerbi SA, Lange TS (2002). All small nuclear RNAs (snRNAs) of the [U4/U6.U5] Tri-snRNP localize to nucleoli; Identification of the nucleolar localization element of U6 snRNA. *Mol Biol Cell* **13:** 3123-37.

Ghetti A, Company M, Abelson J (1995). Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA* **1:** 132-45.

Gietz RD, Woods RA (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350:** 87-96.

Gilbert W (1978). Why genes in pieces? *Nature* **271:** 501.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H *et al* (1996). Life with 6000 genes. *Science* **274:** 546, 563-7.

Golas MM, Sander B, Will CL, Luhrmann R, Stark H (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell* **17:** 869-83.

Gorini G, Ponomareva O, Shores KS, Person MD, Harris RA, Mayfield RD (2010). Dynamin-1 co-associates with native mouse brain BKCa channels: proteomics analysis of synaptic protein complexes. *FEBS Lett* **584:** 845-51.

Gottschalk A, Bartels C, Neubauer G, Luhrmann R, Fabrizio P (2001a). A novel yeast U2 snRNP protein, Snu17p, is required for the first catalytic step of splicing and for progression of spliceosome assembly. *Mol Cell Biol* **21:** 3037-46.

Gottschalk A, Kastner B, Luhrmann R, Fabrizio P (2001b). The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA* **7:** 1554-65.

Gottschalk A, Neubauer G, Banroques J, Mann M, Luhrmann R, Fabrizio P (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J* **18:** 4535-48.

Gozani O, Feld R, Reed R (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* **10:** 233-43.

Grabowski PJ, Padgett RA, Sharp PA (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell* **37:** 415-27.

Granneman S, Kudla G, Petfalski E, Tollervey D (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* **106:** 9613-8.

Guenther UP, Jankowsky E (2009). Helicase multitasking in ribosome assembly. *Mol Cell* **36:** 537-8.

Hartmuth K, Urlaub H, Vornlocher HP, Will CL, Gentzel M, Wilm M *et al* (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A* **99:** 16719-24.

He Y, Andersen GR, Nielsen KH (2010). Structural basis for the function of DEAH helicases. *EMBO Rep* **11:** 180-6.

Hodnett JL, Busch H (1968). Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J Biol Chem* **243:** 6334-42.

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G *et al* (2008). Gene Ontology annotations at SGD: new data sources and annotation methods. *Nucleic Acids Res* **36:** D577-81.

Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78:** 657-68.

James SA, Turner W, Schwer B (2002). How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8:** 1068-77.

Jandrositz A, Guthrie C (1995). Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J* **14:** 820-32.

Jankowsky E, Bowers H (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res* **34:** 4181-8.

Jankowsky E, Gross CH, Shuman S, Pyle AM (2001). Active disruption of an RNA-protein interaction by a DExH/D RNA helicase. *Science* **291:** 121-5.

Jensen KB, Darnell RB (2008). CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol* **488:** 85-98.

Jones MH, Frank DN, Guthrie C (1995). Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc Natl Acad Sci U S A* **92:** 9687-91.

Juneau K, Miranda M, Hillenmeyer ME, Nislow C, Davis RW (2006). Introns regulate RNA and protein abundance in yeast. *Genetics* **174:** 511-8.

Jurica MS, Moore MJ (2002). Capturing splicing complexes to study structure and mechanism. *Methods* **28:** 336-45.

Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Luhrmann R (2006). RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J Mol Biol* **356:** 1248-62.

Kim DH, Rossi JJ (1999). The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA* **5:** 959-71.

Kim SH, Lin RJ (1993). Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. *Proc Natl Acad Sci U S A* **90:** 888-92.

Kim SH, Lin RJ (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* **16:** 6810-9.

Kim SH, Smith J, Claude A, Lin RJ (1992). The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *EMBO J* **11:** 2319-26.

King DS, Beggs JD (1990). Interactions of PRP2 protein with pre-mRNA splicing complexes in Saccharomyces cerevisiae. *Nucleic Acids Res* **18:** 6559-64.

Kiss T (2004). Biogenesis of small nuclear RNPs. *J Cell Sci* **117:** 5949-51.

Kistler AL, Guthrie C (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev* **15:** 42-9.

Konarska MM, Sharp PA (1988). Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc Natl Acad Sci U S A* **85:** 5459-62.

Konarska MM, Vilardell J, Query CC (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell* **21:** 543-53.

Kunkel GR, Maser RL, Calvet JP, Pederson T (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A* **83:** 8575-9.

Kwan SS, Brow DA (2005). The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. *RNA* **11:** 808-20.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al* (2001). Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Lardelli RM, Thompson JX, Yates JR, 3rd, Stevens SW (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16:** 516-28.

Lauber J, Fabrizio P, Teigelkamp S, Lane WS, Hartmann E, Luhrmann R (1996). The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *EMBO J* **15:** 4001-15.

Lebaron S, Froment C, Fromont-Racine M, Rain JC, Monsarrat B, Caizergues-Ferrer M *et al* (2005). The splicing ATPase prp43p is a component of multiple preribosomal particles. *Mol Cell Biol* **25:** 9269-82.

Leeds NB, Small EC, Hiley SL, Hughes TR, Staley JP (2006). The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. *Mol Cell Biol* **26:** 513-22.

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA (1980). Are snRNPs involved in splicing? *Nature* **283:** 220-4.

Lerner MR, Steitz JA (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A* **76:** 5495-9.

Lewis JD, Gorlich D, Mattaj IW (1996). A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. *Nucleic Acids Res* **24:** 3332-6.

Li Z, Brow DA (1993). A rapid assay for quantitative detection of specific RNAs. *Nucleic Acids Res* **21:** 4645-6.

Lin RJ, Lustig AJ, Abelson J (1987). Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. *Genes Dev* **1:** 7-18.

Lin RJ, Newman AJ, Cheng SC, Abelson J (1985). Yeast mRNA splicing in vitro. *J Biol Chem* **260:** 14780-92.

Linder P (2006). Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* **34:** 4168-80.

Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K *et al* (1989). Birth of the D-E-A-D box. *Nature* **337:** 121-2.

Listerman I, Sapra AK, Neugebauer KM (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13:** 815-22.

Liu S, Rauhut R, Vornlocher HP, Luhrmann R (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* **12:** 1418-30.

Liu YC, Chen HC, Wu NY, Cheng SC (2007). A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol Cell Biol* **27:** 5403-13.

Lockhart SR, Rymond BC (1994). Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol Cell Biol* **14:** 3623-33.

Lopez-Bigas N, Audit B, Ouzounis C, Parra G, Guigo R (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* **579:** 1900-3.

Lund E, Dahlberg JE (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science* **255:** 327-30.

Luo HR, Moreau GA, Levin N, Moore MJ (1999). The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5:** 893-908.

Ma X, Yang C, Alexandrov A, Grayhack EJ, Behm-Ansmant I, Yu YT (2005). Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J* **24:** 2403-13.

MacMillan AM, Query CC, Allerson CR, Chen S, Verdine GL, Sharp PA (1994). Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev* **8:** 3008-20.

Madhani HD, Guthrie C (1994). Genetic interactions between the yeast RNA helicase homolog Prp16 and spliceosomal snRNAs identify candidate ligands for the Prp16 RNA-dependent ATPase. *Genetics* **137:** 677-87.

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M *et al* (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298:** 2205-8.

Makarova OV, Makarov EM, Liu S, Vornlocher HP, Luhrmann R (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is

required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J* **21:** 1148-57.

Malca H, Shomron N, Ast G (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol Cell Biol* **23:** 3442-55.

Martin A, Schneider S, Schwer B (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem* **277:** 17743-50.

Martin-Tumasz S, Butcher SE (2009). (1)H, (13)C and (15)N resonance assignments of a ribonucleoprotein complex consisting of Prp24-RRM2 bound to a fragment of U6 RNA. *Biomol NMR Assign* **3:** 227-30.

Martin-Tumasz S, Reiter NJ, Brow DA, Butcher SE (2010). Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. *RNA* **16:** 792-804.

Massenet S, Mougin A, Branlant C (1998). Posttranscriptional modifications in the U small nuclear RNAs. . In: Grosjean H (ed). *Modification and Editing of RNA*. ASM Press: Washington, D.C. pp 201-228.

Massenet S, Pellizzoni L, Paushkin S, Mattaj IW, Dreyfuss G (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol Cell Biol* **22:** 6533-41.

Mayas RM, Maita H, Staley JP (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* **13:** 482-90.

Mayes AE, Verdone L, Legrain P, Beggs JD (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J* **18:** 4321-31.

Meyer M, Vilardell J (2009). The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. *Brief Funct Genomic Proteomic* **8:** 60-7.

Muramatsu M, Hodnett JL, Busch H (1966). Base composition of fractions of nuclear and nucleolar ribonucleic acid obtained by sedimentation and chromatography. *J Biol Chem* **241:** 1544-50.

Nakajima T, Uchida C, Anderson SF, Lee CG, Hurwitz J, Parvin JD *et al* (1997). RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* **90:** 1107-12.

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371:** 297-300.

Neubauer G, Gottschalk A, Fabrizio P, Seraphin B, Luhrmann R, Mann M (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc Natl Acad Sci U S A* **94:** 385-90.

Newby MI, Greenbaum NL (2002). Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol* **9:** 958-65.

Newnham CM, Query CC (2001). The ATP requirement for U2 snRNP addition is linked to the pre-mRNA region 5' to the branch site. *RNA* **7:** 1298-309.

Niggli V, Penniston JT, Carafoli E (1979). Purification of the (Ca2+-Mg2+)-ATPase from human erythrocyte membranes using a calmodulin affinity column. *J Biol Chem* **254:** 9955-8.

Noble SM, Guthrie C (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* **143:** 67-80.

Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R *et al* (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J* **18:** 6119-33.

Nottrott S, Urlaub H, Luhrmann R (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J* **21:** 5527-38.

O'Day CL, Dalbadie-McFarland G, Abelson J (1996). The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem* **271:** 33261-7.

Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101:** 187-98.

Okamura N, Busch H (1965). Base Composition of High Molecular Weight Nuclear Rna of Walker Tumor and Liver of the Rat. *Cancer Res* **25:** 693-7.

Oruganti S, Zhang Y, Li H (2005). Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun* **333:** 550-4.

Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225:** 898-903.

Palacios I, Hetzer M, Adam SA, Mattaj IW (1997). Nuclear import of U snRNPs requires importin beta. *EMBO J* **16:** 6783-92.

Parenteau J, Durand M, Veronneau S, Lacombe AA, Morin G, Guerin V *et al* (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol Biol Cell* **19:** 1932-41.

Parker R, Siliciano PG, Guthrie C (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49:** 229-39.

Patton JR, Jacobson MR, Pederson T (1994). Pseudouridine formation in U2 small nuclear RNA. *Proc Natl Acad Sci U S A* **91:** 3324-8.

Paushkin S, Gubitz AK, Massenet S, Dreyfuss G (2002). The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol* **14:** 305-12.

Perriman R, Barta I, Voeltz GK, Abelson J, Ares M, Jr. (2003). ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc Natl Acad Sci U S A* **100:** 13857-62.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C (2007). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol* **5:** e90.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E *et al* (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24:** 218-29.

Puig O, Gottschalk A, Fabrizio P, Seraphin B (1999). Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev* **13:** 569-80.

Query CC, Konarska MM (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell* **14:** 343-54.

Query CC, Moore MJ, Sharp PA (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* **8:** 587-97.

Rader SD, Guthrie C (2002). A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *Rna* **8:** 1378-92.

Raghunathan PL, Guthrie C (1998a). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* **8:** 847-55.

Raghunathan PL, Guthrie C (1998b). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* **279:** 857-60.

Reyes JL, Kois P, Konforti BB, Konarska MM (1996). The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA* **2:** 213-25.

Riedel N, Wise JA, Swerdlow H, Mak A, Guthrie C (1986). Small nuclear RNAs from Saccharomyces cerevisiae: unexpected diversity in abundance, size, and molecular complexity. *Proc Natl Acad Sci U S A* **83:** 8097-101.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17:** 1030-2.

Rodriguez JR, Pikielny CW, Rosbash M (1984). In vivo characterization of yeast mRNA processing intermediates. *Cell* **39:** 603-10.

Rosbash M, Harris PK, Woolford JL, Jr., Teem JL (1981). The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. *Cell* **24:** 679-86.

Roy J, Kim K, Maddock JR, Anthony JG, Woolford JL, Jr. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* **1:** 375-90.

Ruby SW, Abelson J (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* **242:** 1028-35.

Ruby SW, Chang TH, Abelson J (1993). Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev* **7:** 1909-25.

Ruskin B, Krainer AR, Maniatis T, Green MR (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38:** 317-31.

Sadygov RG, Eng J, Durr E, Saraf A, McDonald H, MacCoss MJ *et al* (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res* **1:** 211-5.

Sapra AK, Khandelia P, Vijayraghavan U (2008). The splicing factor Prp17 interacts with the U2, U5 and U6 snRNPs and associates with the spliceosome pre- and post-catalysis. *Biochem J* **416:** 365-74.

Schaffert N, Hossbach M, Heintzmann R, Achsel T, Luhrmann R (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J* **23:** 3000-9.

Schmitt C, von Kobbe C, Bachi A, Pante N, Rodrigues JP, Boscheron C *et al* (1999). Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J* **18:** 4332-47.

Schneider S, Schwer B (2001). Functional domains of the yeast splicing factor Prp22p. *J Biol Chem* **276:** 21184-91.

Schwer B (2001). A new twist on RNA helicases: DExH/D box proteins as RNPases. *Nat Struct Biol* **8:** 113-6.

Schwer B (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* **30:** 743-54.

Schwer B, Gross CH (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* **17:** 2086-94.

Schwer B, Guthrie C (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349:** 494-9.

Schwer B, Guthrie C (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* **11:** 5033-9.

Seraphin B (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J* **14:** 2089-98.

Seraphin B, Rosbash M (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59:** 349-58.

Shannon KW, Guthrie C (1991). Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev* **5:** 773-85.

Shimba S, Reddy R (1994). Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. *J Biol Chem* **269:** 12419-23.

Siliciano PG, Brow DA, Roiha H, Guthrie C (1987). An essential snRNA from S. cerevisiae has properties predicted for U4, including interaction with a U6-like snRNA. *Cell* **50:** 585-92.

Singh J, Padgett RA (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128-33.

Singh R, Reddy R (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A* **86:** 8280-3.

Sleeman JE, Lamond AI (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol* **9:** 1065-74.

Small EC, Leggett SR, Winans AA, Staley JP (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell* **23:** 389-99.

Smith DJ, Query CC, Konarska MM (2007). trans-splicing to spliceosomal U2 snRNA suggests disruption of branch site-U2 pairing during pre-mRNA splicing. *Mol Cell* **26:** 883-90.

Spiller MP, Boon KL, Reijns MA, Beggs JD (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res* **35:** 923-9.

Spingola M, Grate L, Haussler D, Ares M, Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA* **5:** 221-34.

Staley JP, Guthrie C (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92:** 315-26.

Staley JP, Guthrie C (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* **3:** 55-64.

Stanek D, Rader SD, Klingauf M, Neugebauer KM (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J Cell Biol* **160:** 505-16.

Stevens SW, Abelson J (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci U S A* **96:** 7226-31.

Stevens SW, Abelson J (2002). Yeast pre-mRNA splicing: methods, mechanisms, and machinery. *Methods Enzymol* **351:** 200-20.

Stevens SW, Barta I, Ge HY, Moore RE, Young MK, Lee TD *et al* (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. *RNA* **7:** 1543-53.

Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, Lee TD *et al* (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* **9:** 31-44.

Strauss EJ, Guthrie C (1994). PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro. *Nucleic Acids Res* **22:** 3187-93.

Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ *et al* (2001). The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7:** 382-94.

Tabb DL, McDonald WH, Yates JR, 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1:** 21-6.

Tanaka N, Schwer B (2005). Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22. *Biochemistry* **44:** 9795-803.

Tanaka N, Schwer B (2006). Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. *Biochemistry* **45:** 6510-21.

Tanner NK, Cordin O, Banroques J, Doere M, Linder P (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* **11:** 127-38.

Tarn WY, Hsu CH, Huang KT, Chen HR, Kao HY, Lee KR *et al* (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J* **13:** 2421-31.

Tarn WY, Steitz JA (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* **22:** 132-7.

Trede NS, Medenbach J, Damianov A, Hung LH, Weber GJ, Paw BH *et al* (2007). Network of coregulated spliceosome components revealed by zebrafish mutant in recycling factor p110. *Proc Natl Acad Sci U S A* **104:** 6608-13.

Umen JG, Guthrie C (1995a). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev* **9:** 855-68.

Umen JG, Guthrie C (1995b). Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA* **1:** 584-97.

Umen JG, Guthrie C (1995c). The second catalytic step of pre-mRNA splicing. *RNA* **1:** 869-85.

Venema J, Tollervey D (1999). Ribosome synthesis in Saccharomyces cerevisiae. *Annu Rev Genet* **33:** 261-311.

Vidal VP, Verdone L, Mayes AE, Beggs JD (1999). Characterization of U6 snRNA-protein interactions. *RNA* **5:** 1470-81.

Vidaver RM, Fortner DM, Loos-Austin LS, Brow DA (1999). Multiple functions of Saccharomyces cerevisiae splicing protein Prp24 in U6 RNA structural rearrangements. *Genetics* **153:** 1205-18.

Vijayraghavan U, Company M, Abelson J (1989). Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. *Genes Dev* **3:** 1206-16.

Wahl MC, Will CL, Luhrmann R (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701-18.

Walbott H, Mouffok S, Capeyrou R, Lebaron S, Humbert O, van Tilbeurgh H *et al* (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J*.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C *et al* (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470-6.

Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, Urlaub H *et al* (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16:** 1237-43.

Washburn MP, Wolters D, Yates JR, 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242-7.

Wassarman DA (1993). Psoralen crosslinking of small RNAs in vitro. *Mol Biol Rep* **17:** 143-51.

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A *et al* (2000). A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103:** 457-66.

Weidenhammer EM, Ruiz-Noriega M, Woolford JL, Jr. (1997). Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae. *Mol Cell Biol* **17:** 3580-8.

Weng Y, Czaplinski K, Peltz SW (1996). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol Cell Biol* **16:** 5477-90.

Will CL, Behrens SE, Luhrmann R (1993). Protein composition of mammalian spliceosomal snRNPs. *Mol Biol Rep* **18:** 121-6.

Will CL, Luhrmann R (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13:** 290-301.

Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M *et al* (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* **20:** 4536-46.

Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M, Luhrmann R (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J* **21:** 4978-88.

Wise JA, Tollervey D, Maloney D, Swerdlow H, Dunn EJ, Guthrie C (1983). Yeast contains small nuclear RNAs encoded by single copy genes. *Cell* **35:** 743-51.

Wolin SL, Cedervall T (2002). The La protein. *Annu Rev Biochem* **71:** 375-403.

Xie J, Beickman K, Otte E, Rymond BC (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J* **17:** 2938-46.

Xu D, Nouraini S, Field D, Tang SJ, Friesen JD (1996). An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. *Nature* **381:** 709-13.

Xu Y, Petersen-Bjorn S, Friesen JD (1990). The PRP4 (RNA4) protein of Saccharomyces cerevisiae is associated with the 5' portion of the U4 small nuclear RNA. *Mol Cell Biol* **10:** 1217-25.

Xue D, Rubinson DA, Pannone BK, Yoo CJ, Wolin SL (2000). U snRNP assembly in yeast involves the La protein. *EMBO J* **19:** 1650-60.

Yean SL, Wuenschell G, Termini J, Lin RJ (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408:** 881-4.

Yu YT, Shu MD, Steitz JA (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J* **17:** 5783-95.

Zavanelli MI, Britton JS, Igel AH, Ares M, Jr. (1994). Mutations in an essential U2 small nuclear RNA structure cause cold-sensitive U2 small nuclear ribonucleoprotein function by favoring competing alternative U2 RNA structures. *Mol Cell Biol* **14:** 1689-97.

Zhang D, Rosbash M (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev* **13:** 581-92.

Zhou Z, Licklider LJ, Gygi SP, Reed R (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419:** 182-5.

# Vita

Rea Martine Lardelli was born in 1982 to Diane and Renzo Lardelli in Woodland, California. After graduating from Davis Senior High School, she attended California Polytechnic University, San Luis Obispo, graduating *cum laude* and earning a B.S. in Biochemistry in 2005. She then entered graduate school at University of Texas at Austin in the Department of Biochemistry in 2005.

Permanent address (or email): reamartine@gmail.com

This dissertation was typed by Rea Martine Lardelli.